Adobe® Acrobat® Capture® 3.0

Adobe®

Contents

Chapter 1

Looking at the Work Area

Adobe® Acrobat® Capture® has a main work area and three support work areas.

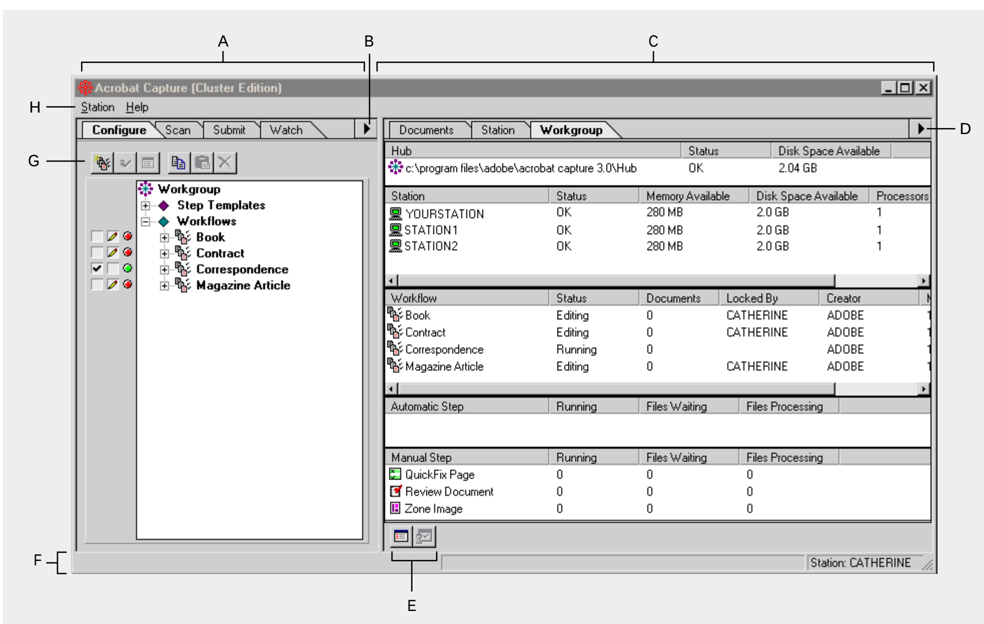
About the Acrobat Capture work areas

Your Acrobat Capture is a suite of four workflow applications, the main application (Acrobat Capture) and three support applications (Zone Tool, QuickFix, and Reviewer) that you start from the main application. You use the Acrobat Capture work area for creating workflows, submitting documents to them, and monitoring the workflows as they process the documents submitted.

A workflow consists of a series of steps, most of which are automatic but some of which are manual. You use the Zone Tool, QuickFix, and Reviewer applications for three kinds of manual steps. For information on the work areas of these applications, see "Looking at the Zone Tool work area" on page 67, "Looking at the QuickFix work area" on page 82, and "Looking at the Reviewer work area" on page 94.

Using the main Acrobat Capture work area

The main Acrobat Capture work area is a window consisting of a left pane and a right pane, with a menu bar above the panes and a status bar below. Each pane has several panels, one shown at a time. All the panels have menus. Some have command bars above or below the main panel area.



A. *Left pane* ***B.*** *Configure panel menu* ***C.*** *Right pane* ***D.*** *Workgroup panel menu* ***E.*** *Workgroup panel command bar*
F. *Status bar* ***G.*** *Configure panel command bar* ***H.*** *Menu bar*

The illustration shows the Configure panel in the left pane and the Workgroup panel in the right. Panel tabs show the other panels available in the left and right panes.

Note: *The illustration shows an Acrobat Capture station's work area after the station has joined a workgroup. If your station has not joined a workgroup, the Configuration and Workgroup panels are empty. For information on joining a workgroup, see* *"Setting up your workgroup station" on page 4.*

Using the left and right panes

You use the left pane for creating workflows and submitting scanned and electronic documents to them. The right pane is for managing and monitoring workflows as they process documents.

To change the relative width of the left and right panes:

Drag the border between the panes to the left or right.

Using panels

Each of the seven tabbed panels in the two panes has its own menu. The Configure, Watch, Documents, and Workgroup panels have command bars for easier access to important menu commands. The Watch, Documents, and Workgroup panels show information in sizable table columns.

Information on using specific panels is in separate places:

- *Configure* panel information is in "Building a workflow" on page 12.
- *Scan* panel information is in "Looking at the Scan panel" on page 27.
- *Submit* panel information is in "Submitting files directly" on page 40.
- *Watch* panel information is in "Setting up watched folders" on page 41.
- *Documents* panel information is in "Looking at the Documents panel" on page 47.
- *Station* panel information is in "Viewing steps and documents in the Station panel" on page 49.
- *Workgroup* panel information is in "Monitoring workgroup performance" on page 50.

To view online help on a panel:

1 Click the panel.

2 Choose Help > Help on Panel.

To choose a command from a panel menu:

1 Position the pointer over the triangle in the upper right corner of the pane containing the panel.

2 Do one of the following to open the menu and choose the command:

- Click the mouse button, and click the command.
- Hold down the mouse button, drag to the command, and release.

You can also open a panel menu with the keyboard shortcuts Alt+1 (left panel) and Alt+2 (right panel).

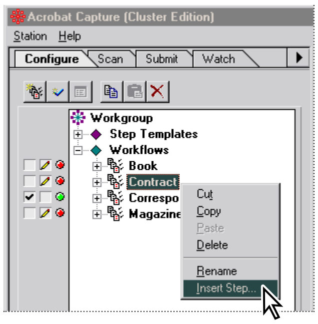
To choose a command from a context menu:

1 Position the pointer over an item in a panel, such as one of the step templates or workflows in the Configure panel.

2 Do one of the following to display the context menu:

• Click the right mouse button.

• Press the context menu key (if available).



3 Click the command.

The Configure, Submit, and Documents panels, and the support applications Zone Tool, QuickFix, and Reviewer have context menus.

Using hierarchies

The contents of the Submit panel and of several Acrobat Capture dialog boxes are Microsoft® Windows® Explorer-style hierarchical representations of your file system. You can expand a parent folder to show its children (subfolders or files) or collapse it to hide them.

The contents of the Configure panel are a similar hierarchical representation of the resources available to your workgroup—step templates and workflows. Expanding the Workgroup icon ( ) shows the step templates and workflows. Expanding a step template shows its properties. Expanding a workflow shows the steps that make it up.

To expand or collapse a folder, or any item in the Configure panel other than Workgroup:

Click the square containing a plus ($\boxed{+}$) or minus ($\boxed{-}$) next to the item's icon.

• When the item is collapsed, the square contains a plus.

• When the item is expanded, the square contains a minus.

If there is no square next to a folder, it cannot be expanded.

Starting Zone Tool, QuickFix, and Reviewer

You start the support applications Zone Tool, QuickFix, and Reviewer from the Manual Step list of the Workgroup panel. You can start a support application only if a workflow that uses its corresponding step is running. See "Running a workflow" on page 37.

To start Zone Tool, QuickFix, or Reviewer:

1 Do one of the following in the Manual Step list of the Workgroup panel:

- To start Zone Tool **,** double-click Zone Image.
- To start QuickFix **,** double-click QuickFix Page.
- To start Reviewer**,** double-click Review Document.

Alternatively, you can select Zone Image, QuickFix Page, or Review Document and choose Run Manual Step from the Workgroup panel menu.

If the value of Files Waiting opposite Zone Image, QuickFix Page, or Review Document is 1 or more, the application starts with the first available file open in it. If Files Waiting is 0, the application starts with no file open it in. It will automatically open the first file that becomes available.

Setting up your workgroup station

Acrobat Capture workgroups are organized as *stations* around a *hub*. A station is a computer running Acrobat Capture. A workgroup can have any number of stations. It can also have only one station (single-station workgroup).

If you are using Acrobat Capture Personal Edition, you are by definition a single-station workgroup.

To identify your workgroup:

Do one of the following:

- Click the Configure tab, and click Workgroup at the top of the panel. If the location of a Hub folder is shown, you are part of the workgroup using that hub.
- Click the Workgroup tab. If the location of a Hub folder is shown at the top of the panel, you are part of the workgroup using that hub.

If no location is shown, you can join an established workgroup from your station. If you plan to use Acrobat Capture Cluster Edition or Acrobat Capture Assistant in an established workgroup and the hub location is on your own computer, you will probably have to relocate to the established workgroup. In either case, before you join, you'll need to know the location of the Hub folder for the workgroup.

Note: *If you are using Personal Edition, the correct location will appear automatically. You can ignore the following section on joining a workgroup.*

Joining a workgroup

If you are part of a workgroup, your station is connecting to the workgroup hub. If you are an Acrobat Capture Cluster Edition or Acrobat Capture Assistant user, your workgroup hub may be located in any folder on any computer accessible by network to all the stations in the workgroup. (It may or may not be located on a station in the workgroup.) If your station isn't part of a workgroup, you need to know the path to the hub to join it.

To join a workgroup:

1 Choose Station > Join Workgroup.

2 Click the browse button (...), and browse to the location of the workgroup hub (for example, \\Server\C\Program Files\Capture\Hub). If your information about the location of the hub is correct but you can't find it, you probably don't have the correct permission for accessing it. Ask your workgroup administrator for help.

3 Click OK.

Checking Station Properties

Properties for an individual station include the automatic steps that can be performed at that station, the location of the station's temporary data folder, and the alerts used at the station. If you are a single-station workgroup, the properties for your own station are set by default and probably don't need changing.

If your station is one of several in a workgroup, the settings for automatic steps will affect the entire workgroup. If a step can't be performed at your station, other stations will need to perform it. If it can be performed at both your station and other stations, Acrobat Capture will determine which station or stations perform the step, balancing the load automatically. For more information, see "Tuning automatic steps" on page 59.

Note: *If you are using Acrobat Capture Assistant, the automatic steps are unselected for your station and cannot be selected. If you submit documents to a workflow containing any of these steps, it will be performed at another station in the workgroup.*

To check station properties:

1 Choose Station > Station Properties.

2 Change settings, as necessary:

- Event sounds are played when a document is finished, when a warning is logged, when a scanner is waiting for action, and when manual work is pending.
- Alerts include flashing taskbar buttons, icons, and panel tabs as well as event sounds.

Note: *If you are using Acrobat Capture Cluster Edition and any of the steps are unselected for your station, make sure you know why they were deselected before you select them.*

3 Click OK.

Looking at workgroup resources

Once you have joined a workgroup, you can examine its resources (workflows, step templates, and stations) in the work area.

Acrobat Capture comes with several sample workflows, and additional workflows may have been built for your workgroup. For information on the sample workflows, see "Using sample workflows" on page 11.

Step templates provide the materials for steps, which are the building blocks of workflows. For information about specific step templates, see "About step templates" on page 7.

To see what workflows and step templates are available:

1 Click the Configure tab.

2 Expand Workgroup, if necessary.

3 Expand Step Templates and Workflows, if necessary. Both individual step templates and workflows are listed alphabetically.

To see the steps that make up a workflow:

1 Expand the workflow. The steps appear in the order they are performed.

2 Expand any step in the workflow to show its properties.

A step in a workflow has the same name as the step template it is based on. It may not have the same properties. For information about the functions of individual steps, see “About individual step templates” on page 8. For information about the properties of individual steps, see “About step properties” on page 18.

To see what other stations are participating in your workgroup:

Click the Workgroup tab.

The stations currently participating in the group, including your own, are listed.

Chapter 2

Building Workflows

Workflows are built from step templates.

About step templates

You build a workflow by inserting steps in it from step templates. Acrobat Capture comes with a set of step templates, and your workgroup may have access to step templates added to the original set. All the available step templates are listed in the Configure panel.

Note: *You can use the Acrobat Capture Software Development Kit (SDK) to build your own step templates, either from scratch or by wrapping existing applications in a step template provided in the SDK.*Most of the Acrobat Capture step templates have configurable properties. See “Configuring Steps” on page 17.

About step template functions

The Acrobat Capture step template set contains 21 templates. See“About individual step templates” on page 8. The 21 templates form six functional groups.

Image preparation steps Includes Convert Image to PDF, Convert Image to TIF, Deskew Image, Filter Image (Kofax), Filter Image (ScanFix), Rotate TIF, Split Multipage PDF, Split Multipage TIF, Uncompress TIF, and Zone Image. Except for Convert Image to PDF, these steps output TIF files that are adapted to optical character recognition (OCR). Convert Image to PDF converts to PDF Image format. The manual step Zone Image uses the application Zone Tool.

Optical character recognition (OCR) step Includes one step, Capture Image. The step’s input is scanned images and TIF files. Its output is one-page Acrobat Capture Page (ACP) files. ACP files are accepted as input by the OCR correction step QuickFix Page, the assembly step Bind Pages, and all three export steps.

OCR correction steps Includes QuickFix Page and Review Document. These manual steps use the applications QuickFix and Reviewer, respectively, to edit individual captured pages and whole captured documents.

Assembly steps Includes Bind Pages and Combine PDFs. These steps assemble pages into multipage documents or documents into multidocument books. Bind Pages accepts ACP files as input and outputs Acrobat Capture Document (ACD) files. ACD files are accepted as input by the OCR correction step Review Document and all three export steps. Combine PDFs uses PDF files as both input and output.

Export steps Includes Export to HTML, Export to PDF, and Export to TXT. These steps accept ACP and ACD files as input. They convert them to files suitable for general external use. To produce permanent files in one of these formats, a workflow needs one of the destination steps following the export step.

Destination steps Includes Archive File, Mail File, and Store File. These steps check in, mail, and store files created in other steps rather than outputting new files. Workflows typically end with a destination step. For information on placing destination steps at other workflow locations, see “Building workflows with parallel steps” on page 15.

About individual step templates

The Acrobat Capture step template set contains 21 templates.

Archive File This destination step accepts files in any format as input and stores them in any document management system (DMS) that is compliant with the Open Document Management API (ODMA). You need to configure this step template to identify the DMS. For configuration information, see “Archive File properties” on page 18. You can find ODMA information at http://www.aiim.org/odma/odma.htm. (The Web site of AIIM, the Association for Information and Image Management International, is http://www.aiim.org.)

Bind Pages This step reassembles multipage documents from one-page files that are the output of Capture Image steps, preserving links, bookmarks, and other features of the original document. The one-page files are in Acrobat Capture Page (ACP) format, the multipage files in Acrobat Capture Document (ACD) format. For the document features preserved in this step, see “Export to PDF” on page 9.

A Bind Pages step also accepts sets of ACD files and outputs a single ACD file that assembles the input files. A workflow designed to create multichapter books that contain cross-chapter links could have two Bind Pages steps, the first (ACP to ACD) to assemble the pages of chapters into whole chapters, the second (ACD to ACD) to assemble chapters into the book.

Capture Image This OCR step accepts images (scanned pages and image files in TIF or other formats) as input and outputs one-page Acrobat Capture Page (ACP) files in which the text found in the image has been recognized and captured. The step also performs *page recognition*—recognition of graphic elements and page layout.

ACP files are useful for review purposes in QuickFix Page steps, but require conversion to formats such as PDF or HTML for regular use. If you capture pages in a variety of languages or fonts, you need to configure this step differently for different workflows or workflow submissions. For more information, see “Configuring Capture Image” on page 19.

Combine PDFs A Combine PDFs step accepts sets of PDF files as input. It outputs a single PDF file that assembles the input files. The step is useful for assembling separate chapter files into book files.

Note: *If you have captured a multipage document as a set of single-page ACP files and want to use an Export to PDF step to turn that document into a PDF document with links, bookmarks, or other advanced features, your workflow should use a Bind Pages step, not a Combine PDFs step, to assemble the pages. Similarly, you should use a second Bind Pages step, not a Combine PDFs step, if you are assembling a multichapter book from several sets of single-page ACP files.*

Convert Image to PDF This image preparation step accepts image files (BMP, JPG, JPEG, PCX, TIF, and PDF Image) as input. It outputs standardized PDF Image files.

Convert Image to TIF This image preparation step accepts image files (BMP, JPG, JPEG, PCX, TIF, and PDF Image) as input. It outputs standardized TIF files.

Deskew Image This image preparation step aligns misaligned images, especially images that have been skewed during scanning. The step is unnecessary in workflows with a Capture Image step. Capture Image has the same functionality built in. However, the Deskew Image step facilitates zoning, so it can be useful before a Zone Image step, even in workflows with a following Capture Image step.

Export to HTML This conversion step accepts ACD or ACP files as input. ACP files are the output of a Capture Image step. ACD files are the output of Bind Pages steps. The Export to HTML step outputs HTML documents consisting of any number of HTML pages. The actual output files are zip files containing the HTML files and image files making up the document.

Processing during the Capture Image step enables Export to HTML output documents to have title, subject, author, and keyword information. (The title becomes the title of the browser window. The other information, placed in the header of the HTML source file, is accessible to spiders and search engines.) However, for this information to be present in the output files, you need to reconfigure the Export to HTML step template or step. See “Configuring Export to HTML” on page 21. Also, the HTML document with this information requires ACD files, not ACP files, as input.

Export to PDF This conversion step accepts ACD or ACP files as input, like Export to HTML. Processing during the Capture Image step that outputs ACP files enables the Export to PDF step to output documents with any of the following PDF features:

- Formatted Text & Graphics or Searchable Image page content (In Acrobat 4.0, PDF Formatted Text & Graphics is called “PDF Normal.” PDF Searchable Image is called “PDF Original Image with Hidden Text.”)
- Bookmarks and links
- Title, subject, author, and keyword information
- Thumbnails
- Structure tags
- Embedded fonts

Formatted Text & Graphics page content, bookmarks, and links are present in the output files by default. For any of the other features to be present, you need to reconfigure the Export to PDF step template or step. See “Configuring Export to PDF” on page 22. Also, for either title, subject, author, and keyword information or bookmarks and links to be present, the PDF document needs to use ACD files, not ACP files, as input.

Export to TXT This conversion step accepts ACD or ACP files as input, like Export to HTML and Export to PDF. It outputs TXT files. For configuration information, see “Export to TXT properties” on page 19.

Filter Image (Kofax) You need a Kofax image processing accelerator supporting ImageControls 3 Gold to run this image preparation step. (You can’t use the step without it.) The step can deshade, deskew, and despeckle images; remove black borders; remove lines and streaks; and filter images. (Filters can fill line breaks; outline regions; and smooth, thicken, or thin characters or lines.) The step works on TIF, JPG, and BMP files, outputting TIF files. For configuration information, see “Configuring Filter Image (Kofax)” on page 23.

Filter Image (ScanFix) You need TMSSequoia ScanFix 4.0 or later to run this image preparation step. (You can’t use the step without it.) The step can remove dot shading zones and stray specks, register pages precisely, remove lines from forms, deskew images, and convert inverse text. It works on monochrome files in TIF, PCX, DCX, and BMP formats, outputting TIF files. For configuration information, see “Filter Image (ScanFix) properties” on page 19.

***Note:** ScanFix cannot process TIF files that use LZW compression. If you need to process such files, use an Uncompress TIF step before the Filter Image (Scanfix) step.*

Mail File This destination step accepts files in any format as input. The step sends a separate e-mail message for each input file, either notifying the recipients that the file has been processed or actually delivering the file as an attachment. You need to configure this step template to provide the recipients, subject line, and text of the message. For more information, see "Configuring Mail File" on page 24.

QuickFix Page This OCR correction step accepts ACP files created with a Capture Image step as input and outputs the same files in edited form. The editing, performed manually with the QuickFix application, is review and correction of text recognized in the Capture Image step. For information on performing the step, see "QuickFixing Word Suspects" on page 81. For information on configuring the step, see "Defining suspects" on page 122.

Review Document This OCR correction step accepts ACD files as input and outputs the same files in edited form. ACD files are the output of Bind Pages steps, which convert sets of ACP files to this format. The editing, performed manually, is review and correction of text and page elements recognized during the Capture Image step. For information on performing the step, see "Reviewing Captured Text and Graphics" on page 93. For information on configuring the step, see "Defining suspects" on page 122.

Rotate TIF This image preparation step rotates the image 180˚, or 90˚ clockwise or counterclockwise. Alternatively, you can use the Capture Image step to rotate images automatically, or use the manual Zone Image step to rotate images one at a time. For configuration information, see "Rotate TIF properties" on page 19.

Split Multipage PDF and Split Multipage TIF These image preparation steps separate multipage image files (PDF Image or TIF) into single-page files of the same format. This is essential in workflows with a Capture Image step that process multipage image files, because Capture Image accepts only single-page files as input.

Store File This destination step accepts files in any format as input and stores them in any accessible folder. The default storage folder is Out, a subfolder of your workgroup's Hub folder. If you use another folder and other stations in your workgroup need access to it, be sure the folder is shared and accessible to all stations. For configuration information, see "Store File properties" on page 19.

Uncompress TIF This image preparation step expands compressed TIF image files. It will be necessary in any workflows using custom steps that require uncompressed TIF files as input. Also, the Capture Image step processes some (but not all) types of compressed TIF files. The Uncompress TIF step must precede a Capture Image step in workflows that process compressed TIF files that the step can't process.

Zone Image This image preparation step accepts images (scanned pages and image files) as input. It outputs a Capture Zone (CZF) file that works with the image file when it is used as input in a following Capture Image step. The CZF file divides the image into content zones (text, images, and so on) to control processing of these contents during the Capture Image step.

The Zone Image step is manual and works a page at a time. As an automatic alternative with documents whose pages have a fixed layout, you can use a CZF file in the Capture Image step itself as a default zone file. For information on performing the Zone Image step, see "Zoning Images for Capture" on page 65. For information on configuring the step, see "Zone Image properties" on page 19.

About online information on steps

The quickest way to get information on the uses of individual steps is to use online help. As an alternative, the About Steps dialog box provides a survey of all the available steps and some background information.

To get information about the available step templates:

Choose Help > About Steps.

To get help on an individual step template or step:

1 Click the Configure tab.

2 Do one of the following:

• Select the step template or step, and choose Help on Step from the Configure panel menu.

• Right-click the step template or step, and choose Help on Step from the context menu.

Online help is available for the original set of Acrobat Capture step templates, and may be available for other available templates.

About workflow design

For complete image conversion, a basic workflow needs to contain at least four steps, in the following order: Capture Image, Bind Pages, any export step, and any destination step. More elaborate workflows have one or more image preparation steps before the Capture Image step, an OCR correction step after Capture Image, or both. Also, workflows can have multiple export or destination steps. For information on multiple steps of the same type, see "Building workflows with parallel steps" on page 15.

If you have custom step templates, they may support workflows devoted to processes other than image conversion. In that case, your design guidelines will depend entirely on the construction of the steps—what file types they accept as input and produce as output.

Using sample workflows

You can use the four Acrobat Capture sample workflows as is, or you can use them as templates, inserting, deleting, and replacing steps as necessary. All of the workflows perform complete image conversion. All contain the following four steps, in this order: Capture Image, Bind Pages, Export to PDF, and Store File. Two of the sample workflows, Book and Correspondence, use only these four steps. The other two add manual steps.

Two of the workflows, Contract and Correspondence, have step properties changed from the default settings. For information on changing properties, see "Configuring Steps" on page 17.

Book

The Book workflow uses default properties for all of its steps. The Export to PDF step has the following default properties:

• Detects and preserves links and structure in the original document.

• PDF page content type is Formatted Text & Graphics. The text can be scaled, indexed, searched, and copied.

• Words falling below a 95% confidence threshold during OCR are treated as suspect, and replaced by the original bitmap in the PDF document.

Contract

The Contract workflow is similar to the Book workflow with two exceptions:

• Contains a manual QuickFix Page step. You use the QuickFix application to perform a quick review and cleanup of text suspects. For information on QuickFix, see “QuickFixing Word Suspects” on page 81.

• After fixing, converts the document to PDF Searchable Image (Exact) rather than PDF Formatted Text & Graphics. PDF Searchable Image documents contain complete bitmap images of the pages of the original document, like the PDF Image documents producible by scanning. But they have recognized text hidden behind the images, and this text can be indexed, searched, and copied like text in a PDF Formatted Text & Graphics document. The format is ideal for cases in which you need to keep the original scanned image of a document for legal or archival purposes but want the document to be searchable.

Note: *In Acrobat 4.0, PDF Formatted Text & Graphics is called “PDF Normal.” PDF Searchable Image is called “PDF Original Image with Hidden Text.”*

Correspondence

The Correspondence workflow is similar to the Book workflow with two exceptions:

• The confidence threshold for suspects is increased to 99% to decrease the chance that an incorrectly recognized word will be included in the PDF document rather than being replaced by the original bitmap.

• Links and structure are not detected, since the intended document is only a page or two.

If you replace the Store File step in this workflow with a Mail File step, you can use the workflow to circulate incoming letters as well as convert them.

Magazine Article

The Magazine Article workflow is similar to the Book workflow, but contains two manual steps, Zone Image and Review Document:

• For each image processed, you use the Zone Tool application before OCR. Zone definition prevents errors in page decomposition (layout analysis) during OCR, or excludes regions of the page, for example, ads, from OCR entirely.

• For each document processed, you use the Reviewer application after OCR. Reviewer provides a complete cleanup of any layout or text errors still remaining.

This workflow is useful for any pages with multiple columns, fancy layouts, graphics that span columns, and so on.

For information on the two manual applications, see “Zoning Images for Capture” on page 65 and “Reviewing Captured Text and Graphics” on page 93.

Building a workflow

Use the Configure panel for building custom workflows, editing existing workflows, and making copies of workflows and steps.

Inserting steps

When you inset steps by dragging step templates, Acrobat Capture guides you by displaying a Prohibited sign (⊘) on top of the step except when you are dragging it over workflows you have locked for editing. When the template is over a locked workflow, it displays (⊘) only when it is over unsuitable locations in the workflow.

To build a workflow:

1 Click the Insert Workflow button ( ), or choose Insert Workflow from the Configure panel menu.

2 Type a name for the workflow in the “Untitled” text box.

3 Insert steps. For instructions, see “Inserting steps in a workflow” on page 13.

4 Click Done.

5 Unlock the workflow so that it can be used. For instructions, see “Running a workflow” on page 37.

Important: *Unlocking a workflow allows any station in the workgroup not only to run it but also to edit it. If a workflow should be edited only at your station, be sure other members of the workgroup understand the restriction before you unlock it.*

Inserting steps in a workflow

Steps in a workflow process input files, in most cases producing output files. Because most steps accept only certain file types as input and also output certain file types, there are restrictions on the way steps can be sequenced.

When you are inserting steps in a workflow, Acrobat Capture enforces these restrictions automatically, limiting your insertion choices to valid ones. For example, if the first step you insert is Capture Image, you can’t insert a Review Document step. This is because Review Document can neither provide Capture Image input nor accept Capture Image output. You can insert a QuickFix Page step after, but not before, the Capture Image step. This is because a QuickFix Page step accepts the output of a Capture Image step, but not the other way around.

Inserting steps by dragging is faster. Using the Insert Steps command gives more precise feedback.

To insert steps by dragging:

1 Drag the template for the first step to the workflow. The first step you drag doesn’t have to be the first step in the sequence of steps.

Acrobat Capture displays a Prohibited sign (⊘) on top of the step except when you are dragging it over workflows you have locked for editing.

2 Drag the template for the next step directly to the correct position in the workflow.

3 Acrobat Capture displays (⊘) when it is over unsuitable locations in the workflow. It displays a horizontal line at suitable drop locations.

4 Continue dragging steps until you have completed the workflow.

To insert steps with the Insert Steps command:

1 Select the workflow.

2 Click the Insert Step button ( ), or choose Insert Step from the Configure panel menu.

3 Choose the first step from the Insert Step box, and click Insert. The first step you choose doesn’t have to be the first step in the sequence of steps.

4 Choose the next step from the Insert Step box.

5 Choose the location of the step from the Where list box, and click Insert.

6 Continue inserting steps, choosing each step from the Insert Step box, choosing its location from the Where list box, and clicking Insert.

If you want to insert a destination step such as Store File, you may be presented with multiple variants of the step. For example, if you have inserted a Capture Image step (which outputs ACP files) and an Export to PDF step, you are shown two Store File steps: Store File (ACP) and Store File (PDF). Be sure to insert the appropriate variant for your workflow. For information on creating workflows with more than one destination step, see “Building workflows with parallel steps” on page 15.

7 Click Done.

Editing a workflow

When you insert a new workflow, it is automatically locked so you can build it. When you need to edit an existing workflow, you need to lock it first.

You can edit any workflow available to your workgroup unless it has been locked at another station. A workflow locked at another station is denoted by a special icon ( ).

Important: *Workflows need to be unlocked before they can run, so you may have access to unlocked workflows that you didn’t build. You should edit such workflows only after obtaining the permission of the builder or workgroup manager.*

To lock a workflow:

Do one of the following:

- Select the workflow, and choose Edit Workflow on the Configure panel menu to activate it.
- Select the edit box for the workflow.

A. *Run box* ***B.*** *Edit box* ***C.*** *Status light*

To rename a workflow:

1 Select the workflow.

2 Choose Rename Workflow from the Configure panel menu.

3 Type a new name to replace the old one.

To delete a workflow or step:

1 Select the workflow or step.

2 Click the Delete button ( ), or choose Delete from the Configure panel menu.

Deleting a step will delete one or more subsequent steps if retaining them would make the workflow invalid. For example, deleting a Capture Image step would delete a subsequent QuickFix Page step, because QuickFix accepts only the output of Capture Image as input.

Copying workflows and steps

If you want to create a workflow that's similar to one you already have, it may be easier to copy and edit the workflow you have than to create the new workflow from scratch. Also, if you need to edit a workflow locked at another station, you can copy it and then edit the copy.

Copying an individual step from an existing workflow to a workflow you are building saves time, especially if the step's property settings are different from those of the corresponding step template and you want to use those different settings.

To copy a workflow or step to the Clipboard:

1 Select the workflow or step.

2 Do one of the following:

- Click the Copy button ( ), or choose Copy from the Configure panel menu.
- Choose Cut from the Configure panel menu.

To insert a workflow from the Clipboard:

1 Select Workflows.

2 Click the Paste button ( ), or choose Paste from the Configure panel menu.

To insert a step from the Clipboard:

1 Expand the workflow you want to paste into.

2 Select a step in the workflow you want to paste after.

3 Click the Paste button ( ), or choose Paste from the Configure panel menu.

Building workflows with parallel steps

A workflow processes a submission as quickly as possible, running steps that are apparently sequential at the same time (in parallel) when it can. For example, a workflow in which a Bind Pages step is followed by an Export to PDF step and then an Export to HTML step, or by two Export to PDF steps with different page-content configurations, runs the two Export steps in parallel. This parallel processing is possible because both Export steps can accept the output of the same Bind Pages step (an ACD file) as input.

Building workflows that do parallel processing is easy. You insert each step in the usual way, after the step whose output you want to process, and let Acrobat Capture handle the fine points of sequencing. However, you might want to examine a parallel workflow someone else has built to understand what it will accomplish.

To examine parallel steps in a workflow:

1 Expand the steps of the workflow.

2 Check the Next Step(s) properties for the steps. A step followed by two parallel steps will show two next steps.

For example, a Bind Pages step followed by parallel Export to PDF and Export to HTML steps will have the Next Step(s) property Export to PDF(1), Export to HTML(1). A Bind Pages step followed by two Export to PDF steps will have the Next Step(s) property Export to PDF(1), Export to PDF(2). In the first example, the numbers 1 and 1 indicate first instances of the two steps. In the second example, 1 and 2 indicate the first and second instances of the same step.

Chapter 3
Configuring Steps

Most step templates and steps are configurable.

About step configuration options

You can configure steps in templates, workflows, and processes.

• You configure a *step template* by resetting its default properties.

• You configure a *step in a workflow* by resetting the properties it has inherited from the step template it used.

• When you customize *processing of a step*, you reset its inherited properties for a particular workflow submission.

When you insert a step in a workflow from a step template, the step inherits the default or reset properties of the template at the time of insertion. If you reset the properties of the step template after the insertion, the properties of the corresponding step stay the same. If you reset the properties of the step after it has been inserted, the properties of the corresponding step template stay the same.

When you customize processing of a step in a workflow, the configuration doesn't apply to other submissions to the workflow or to other instances of the step or the original step template.

Changing step configurations

You change the configuration of a step template or step in the properties dialog box for that template or step. The properties dialog box is the same for the template and its corresponding step.

Note: *The default properties for all the step templates delivered with Acrobat Capture except Archive File and Mail File are usable as is.*

To open a properties dialog box for a step template or a step in a workflow:

1 Make sure the Configure panel is in front.

2 Do one of the following:

• If you are going to configure a step template, expand Step Templates.

• If you are going to configure a step, expand Workflows, and then expand the workflow containing the step.

3 Do one of the following:

• Right-click the step template or step, and choose Properties from the context menu.

• Select the step template or step, and either click the Step Properties button ( ) or choose Step Properties from the Configure panel menu.

When you're configuring a step for a particular submission to a workflow, the dialog box opens automatically during the submission. See "Customizing processing for a submission" on page 40.

To reconfigure properties for a step template or step:

1 Open the properties dialog box.

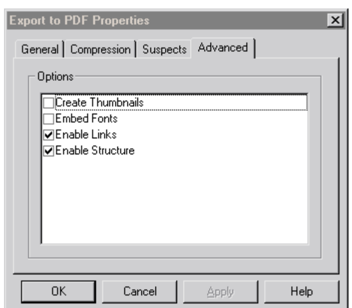
2 Change settings as necessary.

3 Click OK.

To change settings on a properties panel:

1 Change settings on the panel as necessary.

2 Bring other panels to the front to make changes as necessary. (Two of the properties dialog boxes, Export to PDF and Mail File, have more than one panel.) You don't have to click OK to establish settings for individual panels, only for the dialog box as a whole.



Export to PDF Properties dialog box with the Advanced panel in front

About step properties

A number of step templates have configurable properties.

Archive File properties Settings include the document management system (DMS) and file format used in the step. You need to identify the DMS. The file format is based on the filename extension and determined automatically by default. Alternatively, you can select a MIME type from a list of registered types or provide a custom string that you know your DMS can interpret. (You enter the string in the "Other" text box.)

The Archive File Properties dialog box also has a Document Info button. See "Providing document information (metadata)" on page 119.

Capture Image properties Settings include the languages, dictionaries, and font sets used for OCR. For more information, see "Configuring Capture Image" on page 19.

Export to HTML properties There are settings for graphics format, scale, suspects, and advanced HTML features. For more information, see "Configuring Export to HTML" on page 21.

Export to PDF properties There are settings for PDF page content, image compression, suspects, and advanced PDF features. For more information, see "Configuring Export to PDF" on page 22.

Export to TXT properties There is a text encoding setting, with MS-DOS® as the default. The alternatives are Macintosh® and Microsoft Windows.

Filter Image (Kofax) properties You need a Kofax image processing accelerator supporting ImageControls 3 Gold to use this step. You can determine what set of Kofax features to enable and test the set you have established. For more information, see “Configuring Filter Image (Kofax)” on page 23.

Filter Image (ScanFix) properties You need TMSSequoia ScanFix 4.0 or later to use this step. ScanFix settings are stored in a configuration file. For information on the settings, see the ScanFix documentation. In Acrobat Capture, you can identify the path of the configuration file, run ScanFix to edit the file, and choose to log ScanFix messages. By default, messages are not logged.

Mail File properties You need to provide settings for the message that delivers notice of the file or has the file attached. By default, you use the default Messaging API (MAPI) profile to send the file. Alternatively, you can specify a Simple Mail Transfer Protocol (SMTP) server. You also need to provide the recipient, subject line, and text of the message. For more information, see “Configuring Mail File” on page 24.

QuickFix Page properties The settings determine which words are listed as suspects and therefore marked for editing in QuickFix. See “Working with suspects” on page 121.

Review Document properties The settings determine which words are listed as suspects and therefore marked for editing in Reviewer. See “Working with suspects” on page 121.

Rotate TIF properties The default is no rotation. For other ways of setting rotation, see “Configuring Capture Image” on page 19.

Store File properties By default, files are stored in Out (a subfolder of your workgroup’s Hub folder), duplicate files are renamed, and nested input folders are not preserved on output. The default Out folder is described as |Out, with “|” representing the location of the Hub. You can specify any storage folder as an alternative to |Out, including the folder from which the file was submitted.

Zone Image properties By default, new zone files in Zone Tool have a background with no content, composed of a single Excluded zone. You can change the way new zone files define the page area by providing a template zone file for the background. You use this dialog box to specify the path of the template file. See “Using template and default zone files in workflows” on page 75.

The Zone Image Properties dialog box also has a Document Info button. See “Providing document information (metadata)” on page 119.

Configuring Capture Image

By default, the Capture Image step uses the following settings for word recognition:

- A US English character set and spelling dictionary
- A default font set consisting of the Arial, Courier New, and Times New Roman families
- No default zone file
- Automatic page rotation
- Page thumbnail shown in the Steps panel while the page is being processed

About languages

Changing the language from US English changes the character set, the primary spelling dictionary, and (with some languages) the recognition engine. You can select more than one language, in which case there are multiple character sets and multiple language dictionaries. You can also add user dictionaries to the language dictionary or dictionaries, or specify a different font set. See “Working with user dictionaries” on page 122.

About default zone files

You would use a default zone file for zone definition, rather than a preceding Zone Image step, if the pages being captured were identical in layout. See “Using template and default zone files in workflows” on page 75.

About page rotation

You might deselect automatic rotation if you knew that all input pages were properly oriented, or if a step that also handled rotation preceded the Capture Image step. For example, a preceding Rotate TIF step would provide a fixed rotation, and a preceding Zone Image step would allow manual rotation of each image processed.

Creating font sets

When the captured page is displayed, it uses the fonts in the font set specified. Recognition accuracy and page-layout fidelity are optimal when the font set closely matches the original fonts used in the document.

The default font set includes fonts that provide good text-layout fidelity for many common business documents. The set uses one serif proportional font, one sans serif proportional font, and one monospaced font (Times New Roman, Arial, and Courier New, respectively). If you don’t know the original fonts and want to create your own font set, be sure to include an example of each of these three types. You can include any accessible font family that uses the ANSI character set in a font set. For good font recognition, use as few fonts as possible and avoid fonts that are similar to one another. The limit is 16 fonts per set.

To create a font set:

1 Click Edit Sets.

2 Do one of the following:

- Click New to create a new empty font set.
- Select a font set and click Duplicate to create a duplicate font set. If you have several workflows that use the same font set, you might create duplicates, one for each workflow, so that the future revision of a font set required by one workflow wouldn’t affect the other workflows.

3 Enter a name for the set. The text “New Font Set” or “Copy of [selected font set]” is already selected for replacement.

4 Add or subtract font families as necessary. With a new font set you can only add font families. If you created a duplicate to serve as a template for a new set, you may also need to subtract.

5 Click Close.

To add or subtract a font family:

1 Select the font set.

2 Do one of the following:

- To add a font family, select an available font family and click the left arrow (< -).
- To subtract a font family, select the font family in the selected set and click the right arrow (- >).

To delete a font set:

1 Select the set.

2 Press Delete.

Configuring Export to HTML

The Export to HTML Properties dialog box has three panels: General, Suspects, and Advanced.

About general properties (graphics format, downsampling, and scale)

The default graphics format setting is JPEG. The alternatives are CompuServe GIF, interlaced JPEG, and PNG.

The other defaults in the panel are no downsampling of images and 100% scale. The downsampling range is 72 to 300 dots per inch (dpi).

The General panel also has a Document Info button. See “Providing document information (metadata)” on page 119.

About suspect properties (HTML)

An Export to HTML step processes files that have already been processed in a Capture Image step. A word is considered suspect in the Export to HTML step, and a bitmap substituted for it, if confidence that the characters in it have been captured correctly is below a certain threshold or if it has suspect characteristics such as an unknown font or numeric characters. For more information, see “Working with suspects” on page 121.

About advanced properties (HTML)

Processing during the Capture Image step enables the HTML output of an Export to HTML step to have links and a navigation frameset (separate navigation and document panes) enabled. By default, output of the step has both these features.

JavaScript is required for the interactive elements of the frameset to function.

Configuring Export to PDF

The Export to PDF Properties dialog box has four panels: General, Compression, Suspects, and Advanced.

About general properties (PDF page content and size)

The default settings are Formatted Text & Graphics (page content) and Dynamic (page size). The General panel also has a Document Info button. See "Providing document information (metadata)" on page 119.

Page content PDF documents with Formatted Text & Graphics page content contain text that is scalable and can be indexed, searched, and copied to the Clipboard. The documents preserve page formatting and graphic images. PDF Formatted Text & Graphics files are significantly smaller than PDF Searchable Image files, the alternative to the default setting.

PDF Searchable Image documents contain complete bitmaps of the pages of the original document, like the PDF Image documents producible by scanning. But they have recognized text hidden behind the image, so the text can be indexed, searched, and copied like text in a PDF Formatted Text & Graphics document. Use Searchable Image when you need to keep the original scanned image of a document for legal or archival purposes but want the document to be searchable.

Searchable Image (Compact) is a more compressed version of Searchable Image (Exact) in which text regions that contain no color are stored as monochrome rather than full-color images.

Note: *In Acrobat 4.0, PDF Formatted Text & Graphics is called "PDF Normal." PDF Searchable Image is called "PDF Original Image with Hidden Text."*

Page size The default page size, dynamic, is the exact size of the original image. If the original image is a scan, its size may vary slightly from the size of the scanned page itself. If another option is selected (letter, legal, and so on), all varying images are adjusted (snapped) to the single size selected.

About compression properties, including downsampling

The default setting for monochrome is CCITT Group 4 compression and no downsampling. The default for color/grayscale images is Zip compression and no downsampling.

Compression For monochrome compression, the alternative option is Zip compression. CCITT Group 4 is a good general-purpose method. Zip produces the best compression for images that contain repeating patterns.

For color/grayscale compression, the alternative options are JPEG High, Medium, and Low. JPEG compression results in smaller files, so it is the appropriate setting for pages to be viewed on the Web. JPEG is most appropriate for halftone images, such as continuous-tone photographs that contain more detail than can be reproduced on-screen or in print. It is not well suited for the following:

- Images with large areas of a single color, such as screen shots
- Simple images produced with a paint program
- Drawings that have been converted to images

JPEG Low provides the maximum amount of compression, but may result in images with a blocky, quilted look. JPEG High loses such a small amount of detail that most people can't tell that the image has been compressed, but it results in larger files.

Downsampling The downsampling alternatives are 300, 150, and 72 dpi. Downsampling reduces file size by reducing the spatial resolution (pixels per inch) of the image.

About suspect properties (PDF)

An Export to PDF step processes files that have already been processed in a Capture Image step. A word is considered suspect in the Export to PDF step if confidence that the characters in it have been captured correctly is below a certain threshold or if it has suspect characteristics such as an unknown font or numeric characters. If the PDF document exported has Formatted Text & Graphics page content, a bitmap is substituted for the suspect word in the document. For more information, see “Working with suspects” on page 121.

About advanced properties (PDF)

Processing during the Capture Image step gives the PDF output of Export to PDF steps bookmarks, links, thumbnails, or structure, and embeds fonts. By default, the output of the step has links and structure but not thumbnails or embedded fonts, both of which increase file size considerably.

For general information about these PDF features, see the Adobe Acrobat 4.0 Guide.

Links Include standard cross-references (“See page 5”), tables of contents, indexes, chapter titles, and section headings. Titles and headings are listed as bookmarks in a separate navigation panel.

Thumbnails Displayed in the navigation pane along with links, thumbnails are provided for every page in the document. They are disabled by default because they increase file size.

Structure Provides invisible tags for words, paragraphs, table elements, and other elements of a document’s structure. This information can be used by PDF viewers for searching and document navigation. It is particularly useful to visually impaired users using a text-based user interface such as is provided by Adobe Document Server.

Embedded fonts Embedding fonts in a document ensures that the fonts will be the ones you choose, even if those fonts are absent from the computer displaying the document. If you embed fonts, they are limited to the fonts in the font set you have identified in the Capture Image Properties dialog box or assigned in Reviewer. Embedded fonts are disabled by default because they increase file size.

Configuring Filter Image (Kofax)

You need a Kofax image processing accelerator supporting ImageControls 3 Gold to use Filter Image (Kofax).

By default, Filter Image (Kofax) enables only one Kofax feature, Smooth Characters. You can decide which of the Kofax features to enable, or you can let a wizard decide for you. Once a set of features has been decided on, you can perform a before-and-after test on the set from the Filter Image (Kofax) dialog box.

Each Kofax feature has a default configuration and a set of configuration options. The default is appropriate in most cases. For information on these configuration options, see the Kofax documention.

To use a wizard to decide what Kofax features to enable:

1 Click Quick Setup.

2 Select the answer to each of the five questions.

3 Click OK.

To test a Kofax configuration:

1 Click Test.

2 Choose an image file to test, and click Open.

3 Examine the image before and after filtering.

• Scroll and synchronize Before and After images as necessary. Click Sync -> to synchronize the Before image with the After image, or <- Sync to synchronize the After image with the Before image.

• Zoom in or out, as necessary.

To configure an individual Kofax feature:

1 Select the feature.

2 Click Details.

3 Change settings in the options dialog box for the selected feature as necessary.

4 Click OK.

To get information about the software providing a Kofax feature:

1 Select the feature.

2 Click About.

Configuring Mail File

Mail File can be configured for Simple Mail Transfer Protocol (SMTP) or Messaging Application Programming Interface (MAPI). To use MAPI, you need the following:

• MAPI installed and configured on the machine running the Mail File step

• A default MAPI profile (created after MAPI is installed)

• A MAPI-compliant mail client installed and configured on the machine running the Mail File step

For MAPI information, ask your mail administrator, or see your mail client documentation.

When you compose the message, you can insert placeholders for the filename and document name in either the Subject or Message text fields. In a group of one-page files named, for example, Report 001.tif, Report002.tif, and so on, the document name is Report.

To configure the Mail File step for an SMTP server:

1 Click the Sending tab.

2 Select This SMTP Server.

3 Enter the server name, user account, and (optionally) the reply-to address. The user account is not required by all servers.

You need to provide the recipient, subject line, and text of the message. If you have messages with different recipients, subject lines, or text, you'll need to configure this step differently for different workflows or workflow submissions.

To compose the Mail File message:

1 Click the Message tab.

2 Enter the recipients, subject, and message text.

Note: *The recipients will be overridden if MailTo information is included with the document the message concerns. This would happen if a MailTo key is provided as document information (metadata) in any previous step in the workflow, or if the document is accompanied by a JOB file with MailTo information. See "Providing document information (metadata)" on page 119 and "Providing metadata with a JOB file" on page 42.*

3 Select Attach the Current File if necessary.

To send the Mail File message immediately:

Click the Test button on either panel.

After you compose the message, you can test it by trying to send it to the recipients you've specified.

Chapter 4

Scanning with Acrobat Capture

Acrobat Capture works directly with scanners.

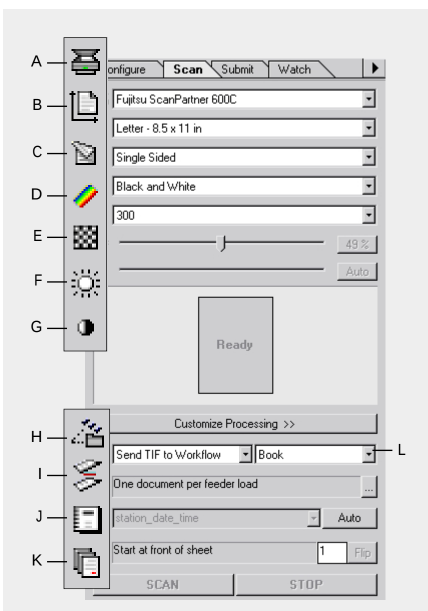
About scanning with Acrobat Capture

Acrobat Capture works with ISIS and TWAIN scanner drivers to produce TIF or PDF image files. Once you have set up Acrobat Capture for use with a scanner, you can use it to scan images to file folders or submit them directly to workflows. For information on scanning directly to workflows, see “Scanning pages to a workflow” on page 39.

You use the Scan panel to set up and use scanners.

Looking at the Scan panel

The Scan panel has controls for selecting and configuring a scanner driver, setting up scanning jobs, and defining the destination of the scanned output.



A. *Scanner driver* ***B.*** *Page size* ***C.*** *Paper source* ***D***. *Color format* ***E.*** *Resolution* ***F.*** *Brightness* ***G.*** *Contrast* ***H.*** *Send*
I. *Separation* ***J.*** *Job name* ***K.*** *Starting sheet* ***L.*** *Target*

Additionally, the Scan panel menu gives access to device-specific driver options, hardware configurations, and commands for saving and loading custom scan settings.

The Customize Processing control on the panel applies only to scans sent to workflows. All other controls and menu items apply to scans sent either to workflows or to folders.

Particular scanner configuration options on the Scan panel and panel menu work only if the options are available for the selected scanner.

Setting up a scanner

You set up a scanner for a particular job or for general purposes. You can save and reuse settings for particular jobs.

Selecting a driver and configuring scanner hardware

Scanner setup for Acrobat Capture begins with selection of a scanner driver. If you have an ISIS scanner, it continues with a check of its hardware configurations. Depending on the scanner, this check can include the I/O port, SCSI device, default paper size, default resolution, and so on.

To select a scanner driver and check hardware configurations:

1 Make sure the scanner is turned on.

2 Select the scanner driver from the scanner driver list ( ).

Note: *If you have an ISIS scanner and it is not on the list, be sure that its driver is in the Windows \Pixtran folder. Most ISIS drivers have installers that place the driver in this folder. ISIS drivers have a .pxn extension.*

3 If a dialog box for the scanner appears, change configurations (as necessary) and click OK.

To change hardware configurations after selection of an ISIS scanner:

1 Choose Configure Hardware from the Scan panel menu. The dialog box that appears is the same one that appeared when you selected the scanner.

2 Change configurations, as necessary. The options available in the dialog box depend on the scanner selected.

3 Click OK.

Changing scanner options

You can change scanner options from Acrobat Capture, either on the upper Scan panel or by opening the scanner's own options dialog box with an Acrobat Capture command. The upper Scan panel is quicker, but the scanner's own dialog box may have controls not found on the Scan panel. In particular, it may have a button for resetting defaults.

Important: *Before you use the upper Scan panel to change options, you should verify that it is working correctly with your scanner.*

To change scanner options in the scanner's options dialog box:

1 With the scanner selected, choose Device Options from the Scan panel menu.

Note: *If you can't choose Device Options (if it is grayed out), obtain the latest driver for your scanner from its manufactuer, install the driver, and try again. However, note that even the latest versions of some TWAIN drivers won't work with Device Options.*

2 Change options, as necessary. The options available in the dialog box depend on the scanner selected.

3 Click OK.

If an option you change also appears in the Scan panel, it will change there as well.

To test an option control on the upper Scan panel:

1 Change the option.

2 Choose Device Options from the Scan panel menu. The corresponding option control in the dialog box should have the same value.

Note: *If you can't choose Device Options (if it is grayed out), obtain the latest driver for your scanner from its manufactuer, install the driver, and try again. If you are still unable to choose Device Options, or if you can use it but its controls and the controls in the Scan panel don't work together, you can't be sure that the panel's controls are working correctly.*

To change scanner options on the upper Scan panel:

With the scanner selected, change options (as necessary) using the following criteria:

• Page size ( ) may not be a standard one. If you select Custom, provide the units, width, and height, and click OK.

• Paper source ( ) options are single sided or double sided. This option describes the pages to be scanned, not whether you have a simplex or duplex scanner. By flipping sheets, you can use a simplex scanner to do double-sided pages.

• Color format ( ) includes black-and-white and grayscale.

• Resolution ( ) is expressed in dots per inch (dpi). For OCR, 300 dpi is usually adequate. If the text is very small (below 8 points), 400 dpi may be better.

• Brightness ( ) can be set manually, with the slider, or automatically. Automatic settings are preferable if available.

• Contrast ( ) can be set manually, with the slider, or automatically. Automatic settings are preferable if they are available.

When a Scan panel option is grayed out, it is unavailable.

To select automatic settings for brightness or contrast:

Do one of the following:

• Make sure the button for brightness ( ) or contrast ( ) is labeled AUTO. If it isn't, click it.

• Check Automatic Brightness or Automatic Contrast on the Scan panel menu.

Saving scan settings

If you can change settings for device options, hardware configuration, or color profile on the Scan panel, you can save the settings for reuse.

To save scan settings:

1 Set up device options, hardware configuration, and a color profile.

2 Choose Save Scan Settings from the Scan panel menu.

3 Provide an identifying filename and click Save. The file is given a *.csc* extension and placed in the Configs subfolder of the Hub folder (|Configs).

To load scan settings:

1 Choose Load Scan Settings from the Scan panel menu.

2 Select a CSC file, and click Open.

Setting up a scanning job

A scanning job is a scan of one or more sheets that use the same scanner settings and job setup. Each page scanned (each side of the sheet) becomes a separate file. Automatic filenaming keeps pages of the same document together. If you use document separation sheets, or if you are scanning fixed length documents, you can scan more than one document in a single job. If you use an automatic document feeder, its capacity does not limit the size of the job.

To set up a scanning job:

1 If necessary, set up the scanner. Even if your scanner is set up, you may have to reset it for a particular job. See “Setting up a scanner” on page 28.

2 Select the method for separating one document from the next. The default method, shown in the separation box ( ), is interactive separation. For more information, see “Separating scanned documents” on page 30.

3 Provide a name for the job in the job name box ( ). The default job name, provided automatically, is made up of the machine name and the date and time of the scan. If the job involves a single document, the job name is also the document name. For more information, see “Naming scanned documents” on page 31.

4 Identify the sheet and (for double-sided scanning) the side the scan starts on. The default starting point, shown next to the starting sheet icon ( ), is the front of sheet 1. For more information, see “Identifying the starting sheet” on page 32.

Separating scanned documents

Separate documents are distinguished by filename. If you are scanning to a folder, you can see the filename groupings in a listing of the files in the folder. If you are scanning directly to a workflow, the groupings take place automatically; in the Documents panel, each group is a single item in the Started and Finished lists, however many files make it up.

The following methods for separating documents are available:

Document separation sheet detection Requires a scanner that supports separator sheets. Using this method, you can scan several documents in one job. The separator sheets mark the boundaries between documents.

One document per feeder load Requires a scanner with an automatic document feeder that can sense when the last sheet in the stack has been fed. If you are using a simplex scanner, you can't use this method to do double-sided scanning.

Interactive Requires you to click a STOP button after the final page in the document has been scanned (default).

Fixed length documents Permits scanning several documents in one job, but only if they are all of the same length. Length is measured by the sheet, and one sheet is one or two pages depending on whether the scan is single sided or double sided.

To change the document separation method:

1 Click the separation button ( ), or choose Document Separation from the Scan panel menu.

2 Select a method. If the method is Fixed length documents, also select the number of sheets.

3 Click OK.

Naming scanned documents

If you stay with the default job name, documents will be easy to identify because the name provides so much information—station, date, and time. However, you may prefer job names shorter than, for example, YOURSTATION_20001007_151215 for a document scanned at YOURSTATION on October 7, 2000 at 3:12 p.m.

Sheet numbers are added to the job name automatically. If the job involves a single document, the job name is also the document name. If the job involves multiple documents, the document names are the job name plus separate document numbers.

Sheet numbers Sheet numbers are added automatically to the job name during the job. If the job name is Report and there are three single-sided sheets scanned to a folder as TIF files in a single document, the files are named Report_00001.tif, Report_00002.tif, and Report_00003.tif. If the job involves multiple documents, the sheet count for each document starts at 00001.

The TIF files for two double-sided sheets (two front pages and two back pages) in a single document would be named Report.00001a.tif, Report.00001b.tif, Report.00002a.tif, and Report.00002b.tif.

Document numbers If the job involves multiple documents, document numbers are added automatically. If the job name is Report and four single-sided sheets are scanned to a folder as two separate two-page documents, the TIF files are named Report_001_00001.tif, Report_001_00002.tif, Report_002_00001.tif, and Report_002_00002.tif.

The document numbers in the example are 001 and 002.

To provide a job name manually:

1 Click the Auto button opposite the job name box ( ) so that it changes to Manual.

2 Enter a job name in the box. The name you enter replaces the automatically provided job name, which is composed of your machine name, the date, and the time.

Identifying the starting sheet

You can start a scanning job at sheet 1, sheet 2, and so on. If you are scanning double-sided sheets, you can start at the back of the starting sheet as well as the front. The default starting point (the front of sheet 1) is shown in the starting sheet box ( ).

To change the starting sheet:

- To change the starting sheet number, enter the new number next to the Flip button.
- To change from front to back, click Flip.

Scanning to a folder

Once you have set up the scanner and the job, you can scan the job to a folder or directly to a workflow. For information on scanning to a workflow, see “Scanning pages to a workflow” on page 39.

To scan a job to a folder:

1 Select Send TIF to Folder or Send PDF to Folder in the send box ( ).

2 Do one of the following:

- Accept the default folder listed in the target box (C:\TEMP).
- Click the browse button ( ), select another folder, and click OK.

3 Do one of the following:

- With an automatic document feeder (ADF), load the stack of sheets, face up or face down depending on your scanner. If the size of the job exceeds the capacity of the ADF, load up to capacity. If you are using separator sheets, place them between documents.
- With a flatbed scanner, load the first sheet face down.

4 Click the SCAN button, or choose the Scan command on the Scan panel menu. This starts the scan at the first page or (with an ADF) the first stack. The button label changes from SCAN to CONTINUE.

5 Complete the job, feeding pages and clicking the CONTINUE, STOP, and Flip buttons, as necessary. For details, see “Scanning with a flatbed scanner” on page 33 and “Scanning with an automatic document feeder (ADF)” on page 33.

Note: *Clicking STOP before the job is completed ends the job and resets Acrobat Capture for another job. You can’t resume a job after you have stopped it. For information on resuming a job interrupted by a scanner problem, see “Recovering from scanner problems” on page 35.*

Scanning with a flatbed scanner

If you are using a flatbed scanner, you can use either the interactive or fixed length method to separate documents. You can scan either single-sided or double-sided sheets.

Note: *If your scanner has an ADF but you are scanning from a flatbed, be sure the ADF isn't loaded when you scan.*

To perform a scanning job with a flatbed scanner:

1 Identify the send method (for example, Send TIF to Folder or Send TIF to Workflow) and target (the path to a folder or the name of a workflow).

2 Place the front of sheet 1 on the scanner.

3 Click SCAN. The label SCAN changes to CONTINUE. While sheet 1 is being scanned, CONTINUE is unavailable (grayed out). When the sheet has been scanned, CONTINUE becomes available and "Start at front of sheet 1" or "Start document 1 at front of sheet 1" changes to "Start at front of sheet 2" or "Start document 1 at front of sheet 2."

4 When the CONTINUE button becomes available, place the front of the next sheet on the scanner and click the button.

5 Repeat the CONTINUE step until you have scanned the last sheet in the job or (with double-sided scanning) the last front side in the job.

6 Do one of the following:

• If you are scanning single-sided sheets, click STOP. Disregard the followings steps.

• If you are scanning double-sided sheets, click the Flip button. With a 20-sheet document, if the starting sheet appears as "Start at front of sheet 21" or "Start document 4 at front of sheet 21" before you click Flip, it reads "Start at back of sheet 20" or "Start document 4 at back of sheet 20" after you click it.

7 Place the back side of the sheet whose front side you just scanned on the scanner, and click CONTINUE.

8 Repeat the CONTINUE step until the job is complete. When you have scanned the back of the sheet whose front side you scanned first, the job ends automatically. You don't have to click STOP.

When a scanning job ends, Acrobat Capture is reset for another job.

Scanning with an automatic document feeder (ADF)

If you are using an ADF, you can use any document separation method. You can scan either single-sided or double-sided sheets.

There are two scanning methods for ADF use:

• The method for performing a single-sided scanning job (single-sided sheets) works when you are using either a simplex or duplex scanner. The method also works for double-sided jobs when you are using a duplex scanner, because there is no need to flip the stack to scan both sides.

• The second method is for performing double-sided scanning jobs on a simplex scanner.

To perform a single-sided scanning job with an ADF scanner:

1 Identify the send method (for example, Send TIF to Folder or Send TIF to Workflow) and target (the path to a folder or the name of a workflow).

2 Be sure the sheets are stacked in the order appropriate for your scanner, ascending (page 1 on top) or descending. If you are using separator sheets, be sure the sheets are placed between documents.

3 Load the stack of sheets into the ADF, face up or face down depending on your scanner. If the size of the stack for the job exceeds the capacity of the feeder, load up to capacity.

4 Click SCAN. Unless your separation method is "one document per feeder load," the label SCAN changes to CONTINUE. (If your method is "one document per feeder load," you have no more steps to perform to complete the job.) While the stack is being scanned, CONTINUE is unavailable (grayed out). When all the sheets in the ADF have been scanned, CONTINUE becomes available.

5 If there are more sheets in the stack, wait till the CONTINUE button becomes available, load the rest of the sheets, and click CONTINUE. Repeat this step as often as necessary.

6 When the CONTINUE button becomes available after you have scanned the last page of the job, click STOP if necessary. Clicking STOP signals the end of the job. With interactive separation, it also indicates the end of the document

When a scanning job ends, Acrobat Capture is reset for another job.

To perform a double-sided scanning job with a simplex ADF scanner:

1 Identify the send method (for example, Send TIF to Folder or Send TIF to Workflow) and target (the path to a folder or the name of a workflow).

2 Be sure the sheets are stacked in the order appropriate for your scanner, ascending (page 1 on top) or descending. If you are using separator sheets, be sure the sheets are placed between documents.

3 Load the stack of sheets into the ADF, front side down if they are in ascending order. If the size of the stack for the job exceeds the capacity of the feeder, load up to capacity.

4 Click SCAN. The label SCAN changes to CONTINUE. While the stack is being scanned, CONTINUE is unavailable (grayed out). When all the sheets in the ADF have been scanned, CONTINUE becomes available.

5 If there are more sheets in the stack, wait till the CONTINUE button becomes available, load the rest of the sheets, and click CONTINUE. Repeat this step as often as necessary.

6 When the CONTINUE button becomes available after you have scanned the last front page of the job, click the Flip button. With a 20-sheet document, if the starting sheet appears as "Start at front of sheet 21" or "Start document 4 at front of sheet 21" before you click Flip, it reads "Start at back of sheet 20" or "Start document 4 at back of sheet 20" after you click it.

7 Flip the stack of sheets and load it into the ADF. The last page (back of the last sheet if the stack is in ascending order) should be the first page scanned. If the size of the stack for the job exceeds the capacity of the feeder, load up to capacity.

8 Click CONTINUE.

9 If there are more sheets in the stack, repeat step 5.

When the scanning job ends, Acrobat Capture is reset for another job. You don't have to click STOP.

Recovering from scanner problems

If your scanner jams or has some other unrecoverable problem while a job is in progress, you can usually resume the job after you have fixed the scanner problem.

To resume a scanning job interrupted by a scanner problem:

1 Don't exit Acrobat Capture or begin a new scanning job.

2 If the job name button is labeled Auto, click it to switch to manual naming.

3 Select the name for the job that was interrupted from the job name box. Even if the interrupted scan used a name provided automatically, it will appear in the box.

Note: *Don't provide a new name. This will make it impossible to resume the interrupted job.*

4 Change the starting sheet number and side to match the sheet number and side of the page following the last page scanned correctly.

5 Click CONTINUE to continue the scan.

Chapter 5
Submitting Documents to a Workflow

Documents are submitted to workflows as groups of scanned pages or one-page files.

About submitting documents

When you submit a document to a workflow, the files comprising it (where each file is one page of the document) are submitted to the first step as *input* and are processed. The *output* of that step then passes to the second step as input, and so on. The final step in a workflow is typically a *destination* step that checks in, stores, or mails a document outputted from a previous step for use outside the workflow. This document is usually a different type and scale from the input document for the first step. For example, a workflow might take a document made up of many one-page TIF files as input to its first step and as its last step store a single multipage PDF Formatted Text & Graphics file based on the TIF files.

Ordinarily, each step in a workflow (for example, Capture Image, Export to PDF, or Store File) is of a different type. Some types of step accept any document type as input. Others accept only certain types. The type of document you can submit to a workflow depends on the type of step the workflow begins with.

A workflow needs to be running before you can submit documents to it. After you submit the document, Acrobat Capture provides a number of tools for monitoring the process. See “Monitoring Workflows” on page 47.

About submission options

You submit a document to a workflow in one of three ways:

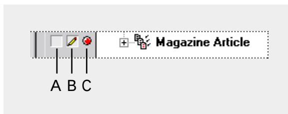
- Scanning its pages to the workflow
- Submitting a group of files containing images of its pages to the workflow
- Setting up a watched folder from which the group of files is submitted to the workflow automatically

Acrobat Capture provides various means of identifying and reassembling the pages of the document.

Running a workflow

A workflow can be running, stopped, stopping, or “editing” (ready to be edited or being edited). If you want to submit documents to a workflow that isn’t running, you need to run it. If it is editing, it is locked. If it was locked at another station, it needs to be unlocked at that station before it can be run. If it was locked at your station, running it unlocks it. Because Acrobat Capture won’t permit you to make a workflow unrunnable (invalid) while you are editing it, you can run it at any time.

A workflow's current status is shown opposite it:



A. Run box B. Edit box C. Status light

To check the status of workflows available to your workgroup:

Do one of the following:

• Click the Workgroup tab. The workflows available are listed in the Workgroup panel, with the status of each: Running, Stopped, Stopping, or Editing. Workflows being edited are locked.

• Click the Configure tab, and expand Workflows. A workflow with a green status light to its left is running.

A yellow status light shows a workflow that is stopping. A red light shows a stopped or locked workflow. If the workflow is locked, a line through the pencil in the edit box to the left of the light shows that it has been locked at another station.

To run a workflow:

1 Click the Configure tab.

2 Expand Workflows in the Configure panel, if necessary.

3 Do one of the following:

• Select the workflow, and choose Run Workflow on the Configure panel menu to activate it.

• Select the run box opposite the workflow.

To stop a running workflow without locking it:

1 Click the Configure tab.

2 Expand Workflows in the Configure panel, if necessary.

3 Do one of the following:

• Select the workflow, and choose Run Workflow on the Configure panel menu to deactivate it.

• Deselect the run box opposite the workflow.

If you stop a running workflow without locking it, any station in the workgroup can run or edit it as needed, and it takes up no system resources in the meantime.

If the workflow is processing documents when you stop it, it will finish the processing before it stops, remaining in a "stopping" state (yellow light) while it does. If you want to terminate this processing, cancel the documents.

To unlock a locked workflow without running it:

1 Click the Configure tab.

2 Expand Workflows in the Configure panel, if necessary.

3 Do one of the following:

• Select the workflow, and choose Edit Workflow on the Configure panel menu to deactivate it.

• Deselect the edit box opposite the workflow.

If you unlock a locked workflow without running it, any station in the workgroup can run or edit it as needed, and it takes up no system resources in the meantime.

Scanning pages to a workflow

When a page is scanned to a workflow, it is converted to a PDF or TIF image that is submitted to the first step in the workflow. From that point, the workflow processes it as it would process any electronic document submitted to it.

To scan pages to a workflow:

1 Make sure the workflow is running.

2 Click the Scan tab to bring the Scan panel to the front.

3 If necessary, set up the scanner and scanner driver for the job. See “Setting up a scanner” on page 28.

4 If necessary, select a new document separation method or starting sheet for the job, or provide a new job name. See “Setting up a scanning job” on page 30.

5 Select Send TIF to Workflow or Send PDF to Workflow in the send box ( ). Use Send TIF to Workflow if the workflow contains a Capture Image OCR step not preceded by a Convert to TIF step.

6 Select the workflow from the list to the right of the send box.

7 Set up customized processing of individual workflow steps, if necessary. See “Customizing processing for a submission” on page 40.

8 Do one of the following:

• With a flatbed scanner, load the first sheet face down.

• With an automatic document feeder (ADF), load the stack of sheets face down. (If the size of the job exceeds the capacity of the ADF, load up to capacity.) If you are using separator sheets, place them between documents.

9 Click the SCAN button, or choose the Scan command on the Scan panel menu. This starts the scan at the first page or (with an AFD) the first stack. The button label changes from SCAN to CONTINUE after you have clicked it.

10 Complete the job, feeding pages and clicking the CONTINUE, STOP, and Flip buttons as necessary. For details, see “Scanning with a flatbed scanner” on page 33 and “Scanning with an automatic document feeder (ADF)” on page 33.

Note: *Clicking STOP before the job is completed ends the job and resets Acrobat Capture for another job. You can’t resume a job after you have stopped it. For information on resuming a job interrupted by a scanner problem, see “Recovering from scanner problems” on page 35.*

Submitting files directly

You can submit any file shown in the hierarchy of folders in the Submit panel to a workflow. The files shown are those acceptable as input by the first step of the workflow selected.

If you treat several files in a submission as a single document, the workflow you submit to needs a Bind Pages step to assemble the document. See “Bind Pages” on page 8.

To submit files to a workflow:

1 Make sure the workflow is running.

2 Click the Submit tab to bring the Submit panel to the front.

3 Select the workflow from the target workflow list opposite the Submit to Workflow button.

4 Select the files to submit in the hierarchy of folders. Selecting a folder selects all the files in the folder. You can select any number of files and folders. Only files that can be submitted to the selected workflow are selectable.

5 Change options, as necessary:

- To remove files on submit, select the box.
- To group files into documents, choose a grouping method from the list above the Submit to Workflow button. See “Creating documents from page files” on page 44.
- To customize processing, display the steps in the selected workflow, and select the steps you want to customize. See “Customizing processing for a submission” on page 40.

6 Click the Submit to Workflow button, or choose the Submit to Workflow command.

Customizing processing for a submission

When you submit scanned pages or electronic documents to a workflow, you can customize individual steps in the workflow. This is a one-submission-only configuration, suitable if you want the step to work differently for this one submission to the workflow. For example, you might want to add document-specific keywords in an Export to PDF step, or you might want to mail each document to a different recipient with a Mail File step.

Customized submissions don’t change the configuration of the customized step permanently. You can customize only the submissions you make from your own station.

For information on all levels of configuration, see “Configuring Steps” on page 17.

To customize processing for a submission:

1 Select the workflow you want to customize.

2 If the arrows on the Customize Processing button point right, click the button to show a list of the steps in the workflow.

3 Select the steps that you want to customize.

4 Click Submit to Workflow.

5 Configure the steps you selected as their properties dialog boxes pop up.

To use the current properties for a step as is, click Cancel in the properties dialog box for the step when the box pops up. The workflow will continue, using the current properties.

Setting up watched folders

In Acrobat Capture, a watched folder is associated with a particular workflow. Whenever a file is added to the folder, it is automatically submitted to the workflow.

About watched folders

When you set up a watched folder and associate it with a particular workflow, Acrobat Capture submits any file to the workflow as soon as it appears in the folder. It deletes the file from the folder upon submission. You can set up multiple watched folders, one for each workflow you use. You can associate multiple workflows with a single watched folder, but only if the initial steps of the workflows have different input types.

A file can be submitted from any computer with access to the watched folder, not just from workgroup stations. When a document consists of multiple files, individual files are submitted as soon as they are placed in the folder. Submissions can be accompanied by JOB files that contain metadata to be used in workflow steps, for example, MailTo information for a Mail File step or keywords for an Export to PDF step. For more information, see “Providing metadata with a JOB file” on page 42.

Working with watched folders

You use the Watch panel to set up a new watched folder, change the properties of an existing watched folder, or remove a watched folder.

To set up a new watched folder:

1 Click the Add Watched Folder button ( ), or choose the Add Watched Folder command.

2 Select a workflow from the Submit Files to Workflow list. It doesn’t need to be running for you to select it.

3 Identify the folder to watch. (Either type the path in the Location field, or click the ( ) button and select the folder in the hierarchy of folders.) The folder can be one already set up with another workflow that processes files of another type.

4 Select Watch Subfolders if you want to include subfolders of the folder you identified.

5 Change options, as necessary:

• Change the method of grouping files into documents. You can treat each file as a document or combine files with the same prefix (document name). For more information, see “Creating documents from page files” on page 44. The workflow you submit to needs a Bind Pages step to assemble the document. See “Bind Pages” on page 8.

• Change the method of signaling that a file placed in the folder is the last file out of a group of files with the same prefix. You can choose a time-out period after the file is placed in the folder, or you can choose an END file. See “Using a time-out period to group files” on page 43 and “Using an END file to group files” on page 43.

6 Click OK.

To change the properties of a watched folder:

1 Do one of the following:

• Double-click the folder.

• Select the folder, and either click the Watched Folder Properties button ( ) or choose the Watched Folder Properties command.

2 Make the changes. The options are the same as when you set up a watched folder. For more information, see “Setting up watched folders” on page 41.

3 Click OK.

To monitor a watched folder:

Check the folder’s icon:

• A closed eye ( ) means the associated workflow is stopped.

• An open eye ( ) means the associated workflow is running.

• An exclamation point ( ) means there is a problem with the associated workflow.

To remove a watched folder:

1 Select the watched folder from the folder list.

2 Click the Remove Watched Folder button, or choose the Remove Watched Folder command.

Submitting files to watched folders

You can use JOB files to add metadata to files submitted to watched folders, and you can control the way the files are grouped into documents with time-outs and END files.

Providing metadata with a JOB file

When you submit files to a watched folder, you can submit a JOB file along with them. The JOB file contains metadata used in certain steps of the workflow associated with the watched folder.

About JOB files A JOB file takes the form of a Windows INI file, and can use any metadata keys you have defined. For example:

```
[Metadata]

Author=Jean Lee

MailTo=jlee@adobe.com

Keywords=capture,workflow,workgroup
```

The JOB file has the name of the document plus the extension *.job*. For example, the JOB file Report.tif.job might accompany the files Report1.tif, Report2.tif, and Report3.tif.

About JOB file metadata When you place a JOB file in a watched folder along with a document and the document is submitted to a workflow, the metadata keys in the JOB file are available to the steps in the workflow. For example, the MailTo key value would provide recipient information in a Mail File step, or the Keywords key value would provide keyword information for an Export to PDF step. JOB file metadata overrides any configuration data already entered for the step.

To provide a JOB file for a document submitted to a workflow from a watched folder:

1 Create a JOB file for the document. The file should be an unformatted text file.

2 Place the JOB file in the watched folder. After the first file in the document is submitted to the workflow, the JOB file will be deleted automatically. (The metadata in it is moved to a database maintained while the document is being processed.)

3 Place the files making up the document in the watched folder. Be sure the JOB file is in the folder before you place the first file.

Using a time-out period to group files

If you are grouping files with the same prefix as documents, and if you set the time-out period at 15 seconds, files with the same prefix submitted at intervals of less than 15 seconds are considered part of the same document. The clock returns to 0 after each file is submitted. The last file submitted before the time-out period is exceeded is treated as the last file in the document.

In the 15-second example, the time-out period stays at 15 seconds while the files making up a document are submitted. In practice, however, time-out logic adapts to the rate at which files appear as follows:

• The time-out period will increase automatically, as necessary, so that it is always at least three times as long as the longest interval between file submissions to the document being submitted.

• If the time-out period increases, it reverts to the original setting (15 seconds in the example) after the last submission to the document.

Using an END file to group files

If the files in a document are being submitted to a watched folder over a congested network, you can't be sure that a time-out will occur only after the last file has been submitted. Similarly, if the pages of the document are being generated one at a time and the delay between pages is hard to predict, the time-out period is hard to establish in advance.

In either case, the alternative is to place an END file in the folder after placing the last page there. The END file has the name of the document plus the extension *.end*. For example, the END file Report.tif.end might be the last file in a submission that included Report1.tif, Report2.tif, Report3.tif, and Report.tif.job.

To signal the last page of a document with an END file:

1 Be sure End of Document Detection in the Watched Folder Properties dialog box is set to END file. For instructions, see "To change the properties of a watched folder:" on page 42.

2 Create an empty END file for the document.

3 Place all the files making up the document in the folder.

4 Place the END file in the folder. After all the files making up the document plus the END file are submitted to the workflow and deleted from the watched folder, the END file is automatically deleted.

Creating documents from page files

Acrobat Capture treats all files submitted to workflows as documents or parts of documents. (This includes pages scanned to workflows that become one-page files after scanning.) Submissions are either documents consisting of single files or documents consisting of groups of files. For example, a single book is scanned to a workflow as many separate pages, and each page enters the workflow as a separate file, but the whole group of pages is a single document.

About file grouping methods

Prior to submission, you identify the group of files making up a document by giving the document a name. How you can identify the group depends on how you submit the files:

- If you scan pages to a workflow, the grouping depends on the job name you provide for the scanning job. With a multidocument job, a document number is added to the job name automatically. For more information, see "Naming scanned documents" on page 31.
- If you submit files directly to a workflow, you can treat each file you submit as a document, or you can combine files with the same prefix, files in the same folder, or all files you have selected. For information, see "About prefixes" on page 44.
- If files are submitted automatically from a watched folder, you can treat each file you submit as a document, or you can combine files with the same prefix.

About prefixes

A prefix is a group of characters *p* that appears first in one of the following patterns:

pn[l].x

p_n[l].x

p-n[l].x

p n[l].x

where *p* is any string of characters beginning with an alphabetic character, *n* is one or more numeric characters, *l* is an optional letter a or b , and *x* is some valid extension. For example, mydocument001.tif, mydocument_001.tif, mydocument-001.tif, mydocument 001.tif, mydocument001a.tif, mydocument001b.tif, and so on.

If you have scanned a group of pages to a folder and are now submitting the files for the pages to a workflow, you can use the "same prefix" method. The job name for a one-document scanning job, or the job name plus document number for a multidocument job, will serve as the prefix or prefixes.

Grouping files

You use the same grouping method when you submit files to a workflow directly and when you submit them to a watched folder.

To group files into documents:

1 Choose one of the following:

- Treat each file as a document
- Treat files with the same prefix as one document. See “About prefixes” on page 44.
- Treat files in same folder as one document (not from watched folders)
- Treat all checked files as one document (not from watched folders)

2 If you chose “all checked files,” provide a name for the document, and click OK.

If you choose “each file,” the document name will be the same as the filename. If you choose “same prefix” or “same folder,” the document name will be the prefix or the name of the folder.

Chapter 6

Monitoring Workflows

Workflows process documents automatically, but they should be monitored.

About workflow monitoring

If you submit pages or page files to a workflow containing manual steps, you or someone else will need to perform the steps. Otherwise, there's usually nothing to do with a workflow after the submission. The workflow does all the work.

However, for troubleshooting purposes, Acrobat Capture provides three means of monitoring files as they are processed by workflows—the Documents panel, Station panel, and Workgroup panel.

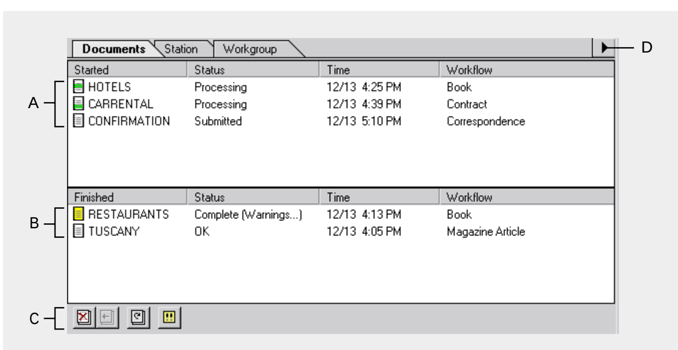
For information on monitoring watched folders, see "Working with watched folders" on page 41.

Monitoring document processing

The Documents panel tracks documents from the time they are submitted to a workflow to the time they are finished.

Looking at the Documents panel

The Documents panel has an upper section listing documents for which processing has started and a lower section listing documents for which processing is finished. The panel menu has commands for working with individual documents and monitoring document events. These same commands are in the command bar at the bottom.



A. *Started list* ***B.*** *Finished list* ***C.*** *Command bar* ***D.*** *Document panel menu*

Working with the Started list

A document appears in the Started list as soon as the first of the files that comprise it is submitted to a workflow. It remains in the list until the workflow has processed all the files that comprise it. If you are in a multistation workgroup, any document submitted to any workflow in the workgroup appears in the list, whether or not it was submitted from your station.

Started document status The status of each started document is listed, together with a color-coded document icon:

- White for *Submitting* and *Submitted*
- An animated green stripe for *Processing* or for *Processing (Warnings ...)*
- Red for *Hung (Warnings ...)*

Warnings A document can continue to be processed even if one or more of its pages can't be processed, providing a placeholder page in these cases. Such documents have the status *Processing (Warnings ...)* in the Started list. In the Finished list, they appear with either a yellow warning (complete) or a red warning (failed), and a log provides further information about them.

If a document is hung, the reason may be low disk space, a network outage, or corrupted files. Acrobat Capture will retry hung documents. Alternatively, you can retry a hung document at any time. You can also cancel a document at any time.

To retry document processing:

1 Select a document in the Started list. If the document is hung, the Retry command will be enabled.

2 Do one of the following:

- Choose Retry Document from the Documents panel menu.
- Click the Retry Document button ( ) in the command bar.

To cancel document processing:

1 Select a document in the Started list.

2 Do one of the following:

- Choose Cancel Document from the Documents panel menu.
- Click the Cancel Document button ( ) in the command bar.

Working with the Finished list

A document appears in the Finished list (accompanied by a "Done!" sound effect) when workflow processing finishes, whether or not it was successful.

Finished document status The status of each finished document is also listed, with a color-coded document icon:

- White for *OK*
- Yellow for *Complete (Warnings ...)*
- Red for *Failed (Warnings ...)*

Warnings For warnings, the Document panel tab flashes red, and there is an “Error!” sound effect.

You can get further information about a warning state by viewing the document event log. See “Using event logs” on page 53. A yellow warning usually means that a page couldn’t be processed and has been replaced by a placeholder page.

To clear documents from the Finished list:

Do one of the following:

- Choose Clear Finished Documents in the Documents panel menu.
- Click the Clear Finished Documents button ( ) in the command bar.

Clearing documents simply removes their names from the list. It does not delete the document files.

Monitoring steps performed at your station

The Station panel shows you the steps running at your own station. A step will run at your station only if it is enabled for performance in your Station Properties dialog box and a workflow containing it is running.

Steps in the Station panel are black and white when they are idle and colored when they are processing files.

- Automatic steps that are processing files are animated.
- Manual steps have a blue border. If the application corresponding to a manual step is minimized or otherwise inactive, clicking within the border activates it.

Viewing the Station panel in different versions of Acrobat Capture

Depending on the version of Acrobat Capture you are using, the Station panel provides a different view of automatic steps:

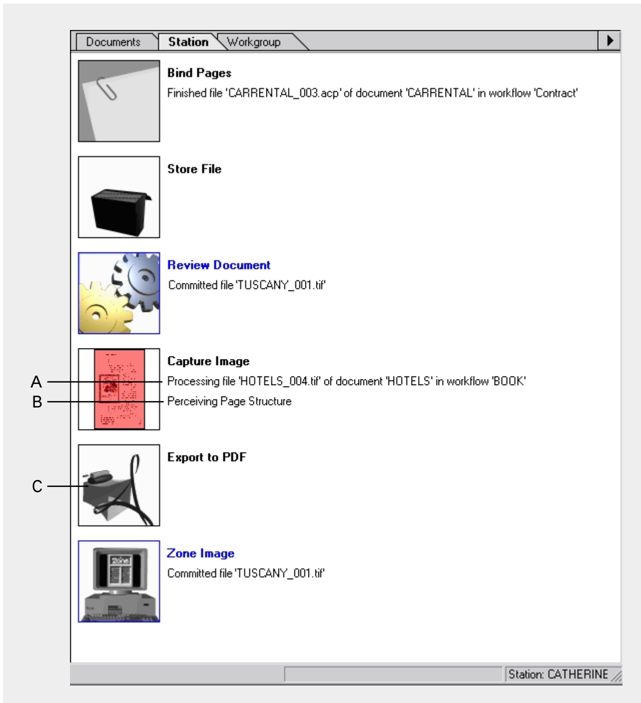
- *Personal Edition* shows all the steps being performed in all the workflows running.
- *Cluster Edition* shows only the steps being performed at your own station. Because workflows will typically use all the stations in your workgroup to process documents, your own Station panel will probably give you only a partial view of any workflow in action. The Workgroup panel provides the complete view.
- *Capture Assistant* shows only manual steps. (It does not perform automatic steps.)

Viewing steps and documents in the Station panel

The Station panel provides detailed information about steps and documents that supplements information in the Workgroup and Documents panels.

Steps For a step that has several phases, the Station panel shows each phase as it happens. (The Workgroup panel shows the number of files waiting to be processed by the step or the performance of the step at other workstations.)

Documents For a document made up of several files, the Station panel shows the step processing each file in turn. (The Document panel shows the status of the document as a whole.)



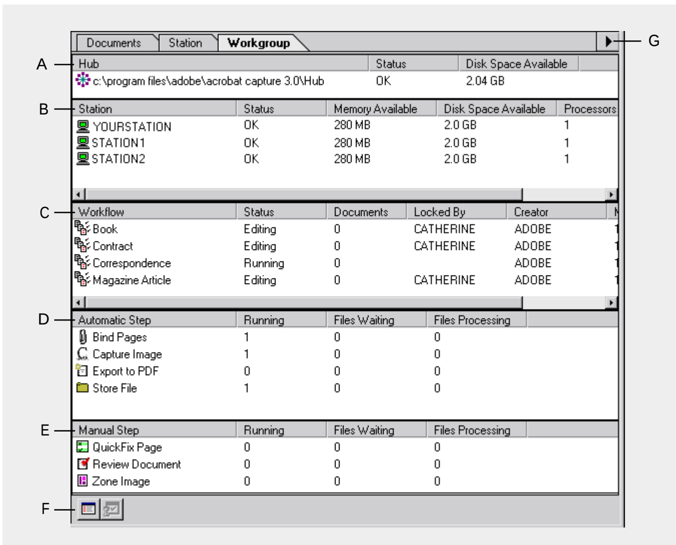
A. *File being processed by the step* ***B.*** *Current phase of the process* ***C.*** *Inactive step*

Monitoring workgroup performance

The Workgroup panel provides status information about your workgroup's hub, the stations joined to it, and the workflows and steps the workgroup is currently running.

Looking at the Workgroup panel

The Workgroup panel has a menu with commands for monitoring station events and running manual steps. These same commands are in the command bar at the bottom.



__A.__ Hub __B.__ Station __C.__ Workflow __D.__ Automatic step __E.__ Manual step __F.__ Command bar __G.__ Workgroup panel menu

The panel tab flashes red when an error message is added to the station event log or when there's work waiting for a manual step. When there's work waiting, the icon of the manual step flashes as well.

Using the Workgroup panel

If you are a one-station workgroup using Acrobat Capture Personal Edition, the Workgroup panel is the most comprehensive monitor of your station's performance.

If your station is part of a multistation workgroup, the Workgroup panel monitors the entire group. If you are managing such a group, see "Monitoring the Workgroup panel (multistation workgroups)" on page 63 for additional information.

Hub In most cases the hub drive should have at least a 100 MB cushion of space when Acrobat Capture is operating at full load (that is, when a typical number of pages is in the pipeline). If this cushion proves inadequate, you can calculate how much disk space should be available when Acrobat Capture is idle by using the formula

<space used> times 3

where <space used> is the disk space used when you are at full load. For example, if you know that about 50 MB are used when you have 25 pages in the pipeline and 25 pages is your typical flow, you should have at least 150 MB available when Acrobat Capture is idle.

For more information, see “Calculating hub drive space needs” on page 61.

Important: *Unless you’re sure about what you’re doing, don’t delete files from the Hub folder or from any of its subfolders except Out. Doing so can cause processing errors. The Out subfolder is the default location for files stored with the Store File step, so it can accumulate files that need deleting. You should avoid using this folder as a permanent file destination.*

Station The status of each station in the workgroup station is listed, with a color-coded station icon: Green for *OK* and Red for *Warnings.*

For warnings, the Workgroup panel tab flashes red, and there is an “Error!” sound effect,

When operating at full load, a station should ordinarily have at least 50 MB of memory (physical memory) available. Or the physical memory available should be something more than the total memory, physical and virtual, allocated to a typical workload for the station. For more information, see“Calculating memory needed for additional processors” on page 62.

Similarly, at full load each station should ordinarily have at least 100 MB of disk space available. There are two exceptions:

- If the station uses multiple processors, you need to calculate the space required more carefully. See “Monitoring the Workgroup panel (multistation workgroups)” on page 63.
- If you are using Acrobat Capture Personal Edition, the hub drive and the station drive are the same. The space needed on the station drive is unlikely to be greater than the space needed on the hub drive.

Workflow The Workflow list shows all the workflows available to your workgroup and indicates the current state of each one: running, stopped, or being edited. (This information is also available in the Configure panel.) The list also indicates the current load of documents on each workflow.

Automatic step With a single station using a single processor, the number of instances of any step running will always be 1, and the number of instances of files processing for that step will never be more than 1. With multistation workgroups or multiprocessor stations, the number of instances may be greater.

The number of files waiting for a step gives you a picture of the load on the various steps. For information on load balancing with Acrobat Capture Cluster Edition, see “Tuning automatic steps” on page 59.

Manual step When work is waiting for a manual step, its icon flashes, the Workgroup panel tab flashes red, and an “Available!” sound effect plays.

With a single station, the number of instances of any step running will be 0 or 1. The number of instances of files processing for a running step will never be more than 1. With multistation workgroups, the number of instances of running steps or files processing may be greater than 1.

With either single-station or multistation workgroups, the count of files waiting can be any number. If you like to work with Zone Tool, QuickFix, or Reviewer in sessions of some fixed length, you can use this count to determine when to begin a session.

Using event logs

If the status of a started or finished document is shown as a warning in the Documents panel, the document event log provides the warning message.

If the status of a station is shown as a red warning in the Workgroup panel, the station event log provides the warning message. Station event messages typically concern processing or resource allocation failures.

Some warnings refer to agents. An *agent* is a program that performs one or more steps. For example, the Capture agent performs the Capture Image and Deskew Image steps.

To examine the document event log:

1 Click the Documents tab.

- Double-click a document in the Started or Finished list.
- Select a document in the Started or Finished list, and either click the Document Event Log button ( ) in the command bar or choose Document Event Log in the Documents panel menu.

2 If necessary, choose an option:

- Click Refresh to see any events that may have occurred since you opened the log.
- Click Save As, and save the log. The log is in HTML format. You can view it with any Web browser.

3 Close the log.

To examine the station event log:

1 Click the Workgroup tab.

2 Do one of the following:

- Double-click a station in the Station list.
- Select a station in the Station list, and either click the Station Event Log button ( ) in the command bar or choose Station Event Log in the Workgroup panel menu.

3 If necessary, choose an option:

- Click Refresh to see any events that may have occurred since you opened the log.
- Click Save As, and save the log. The log is in HTML format. You can view it with any Web browser.

4 Close the log.

Chapter 7

Managing Workgroups

This chapter is particularly for the person who plans, sets up, or leads a multistation Acrobat Capture workgroup. Parts of it will be useful to the person who operates a single-station workgroup, especially a multiprocessor station or one with a large throughput.

About the workgroup manager role

Any multistation workgroup needs to be planned and set up, and once it's in operation it needs to be monitored and tuned. Also, someone has to be identified as the main information source, troubleshooter, arbitrator of resource conflicts, and so on. The workgroup manager role defined in this chapter is simply the sum of all the tasks that fill these needs. The role can be filled by one person, or the tasks can be shared among several people.

The hub of an Acrobat Capture workgroup is a folder on a drive accessible to the stations of the workgroup. Each station in the group needs read and write access to the hub folder to operate. To provide this access, the manager needs administrator privileges on the hub drive.

Note: *In a multistation workgroup, at least one of the stations is using Acrobat Capture Cluster Edition and each of the other stations is using either Cluster Edition or Acrobat Capture Assistant.*

Visualizing your workgroup

If you're beginning an Acrobat Capture workgroup from scratch, it helps to try to visualize the thing you're building. Technically, the workgroup is a cluster of stations around a hub. The stations could be computers anywhere on a network, and the computer with the hub folder needs only to be accessible, not centrally located. For simplicity, however, it helps to imagine a number of computer workstations in a single room, with a hub computer in the center:

As workgroup manager, you are at one of the stations in the group. You've done the following, more or less in order:

- Set up your station, and established the hub.
- Built some workflows for your group's projects.
- Arranged for the scanner or scanners your workgroup will be using to be connected to stations in the workgroup and to be set up for Acrobat Capture use.
- Identified the other stations in the workgroup, and arranged for Acrobat Capture Cluster Edition or Acrobat Capture Assistant to be installed on each of them.
- Made sure that these stations have actually joined the workgroup—that is, they are connected to the hub.
- Organized the team of people who will perform manual steps on stations in the workgroup.
- Optionally, set up one or more watched folders for submissions to workflows.

At this point, workflows can begin processing. Technically, scans and submissions to workflows can come from any station in the workgroup at any time and volume, and submissions can also come from watched folders potentially accessible to a wider circle of users. Practically, you have probably established guidelines for the volume, content, and timing of scans and submissions, and exercise more or less control over the process.

You have probably also set up roles and schedules for the people who perform manual steps. Workgroup members have to know who is to perform the different types of manual steps, and their work schedules need to be coordinated with the flow of documents through workflows containing those steps.

Finally, you are continuously monitoring the workgroup's practices, resources, and workflows to fit them to the objectives of the group.

In this picture, it's a given that you exercise control over workgroup resources and participants.What's up to you is how you exercise the control—by doing much of the work yourself, providing and enforcing guidelines, or building a smoothly functioning workgroup team.

Planning your workgroup's setup

Setup planning involves answering several questions:

- How many stations should there be?
- Should the stations form a single workgroup or multiple workgroups?
- Where should the workgroup hub or hubs be located?
- Which versions of Acrobat Capture should be installed on which stations?
- How should memory, disk space and performance, and processing power be distributed among the stations?

The questions are open-ended, in the sense that your workgroup can easily be modified after you have set it up. For that reason, they are best answered with a simple workgroup already set up and appropriate workflows built for it.

To plan a workgroup:

1 Set up a simple workgroup, perhaps just a single Acrobat Capture Cluster Edition station.

2 Submit a typical job to a workflow:

- Use the type of document you will normally submit, for example, a two-column technical manual with four-color illustrations some of which straddle the columns. Color increases the processing load considerably, so it's important that the color content of the documents be typical.
- Submit roughly the number of pages you will submit in a typical job.
- Use a workflow you will actually use. Be sure it includes any manual steps that will be part of your actual workflow. Be sure to configure important steps, such as Capture Image and Export to PDF, as they will actually be configured.
- Use the submission method you will use—scanning, submitting directly, or submitting from a watched folder.

3 Collect throughput (page/minute) data on the submission for the following:

- The workflow
- Time-consuming steps such as Capture Image
- Each manual step
- Scanning (if you are scanning to the workflow)

For comparison, a single station with a single 200 MHz processor running the Book workflow has a throughput of about 4 pages/minute with documents submitted from the file system (that is, not scanned). The sample Book workflow has a Capture Image step, an Export to PDF step, and no manual steps. The PDF documents are exported with Formatted Text & Graphics page content.

4 Consider your total projected volume of submissions, and decide how much (if any) this throughput needs to be increased.

5 Consider the variety of ways of increasing throughput. See “Increasing throughput” on page 57.

6 Consider other limiting and enabling factors. See “Considering limiting and enabling factors” on page 60.

7 Develop a plan based on these considerations.

Increasing throughput

You can increase throughput in the following ways:

- Add Acrobat Capture Cluster Edition stations to the workgroup, use faster processors at existing stations, or convert single-processor stations to multiprocessor stations. For more information, see “Adding processing power” on page 58.
- Improve the efficiency of your workflow’s time-consuming OCR step, Capture Image, by having appropriate image preparation steps precede it. If you don’t have the software to perform the Filter Image (Kofax) or Filter Image (ScanFix) step, consider obtaining it. For more information, see “About step template functions” on page 7.
- Increase the number of hours per day devoted to manual steps. See “Adding manual-step resources” on page 58.
- Add scanning throughput—faster scanners, more scanners, or more efficient scanner use. If you have several scanners to use, consider dedicating them to different input types—large format, color, monochrome, and so on. This saves time you would otherwise spend reconfiguring them for the different types.
- Make the hub more efficient. See “Locating the hub for best performance” on page 58.
- Set up multiple small workgroup clusters, with multiple hubs, rather than one large cluster. See “Setting up multiple workgroups” on page 59.
- Tune the steps in your workflows to the available stations and processors, for example, using a fast processor to work exclusively on a time-consuming automatic step. For more information, see “Tuning automatic steps” on page 59.

• Distribute document submissions to different workflows so that no one workflow becomes a bottleneck. For more information, see “Keeping automatic steps busy” on page 60.

Adding processing power

Unless your main bottleneck is in a manual step, adding processing power is the most effective and predictable way of increasing throughput.

If an Acrobat Capture Cluster Edition station with a single 200 MHz processor has a throughput of 4 pages/minute, the throughput will roughly double to 8 pages/minute in the following circumstances:

• The processor is replaced by a 400 MHz processor.

• The station adds a second 200 MHz processor.

All things being equal, this rough throughput/processor-speed ratio will hold for any processor speed and for any number of processors per station. For more information, see “Calculating memory needed for additional processors” on page 62.

If you make a processor-speed change—for example, replace your processor with one twice as fast—and the ratio doesn’t hold up, you need to look elsewhere for ways to increase throughput.

If a second Cluster Edition station with a 200 MHz processor is added to the workgroup, throughput will come close to doubling. (The improvement will be slightly less than the improvement from adding a second 200 MHz processor to the first station.) The ratio will hold for any increase in the number of Cluster Edition stations in the workgroup, but only up to a point. For any workgroup, that point is determined by the bandwidth of the network it uses—the speed of the network and the level of background network traffic.

Adding manual-step resources

If your throughput is slowed by a manual-step bottleneck, you can improve it by increasing the amount of operator time devoted to performing the step or by making the performance more efficient:

• Adapt the scheduling of workflows using manual steps to the schedules of the workgroup members who perform the steps.

• Add stations so that several people can work on the same manual step at the same time. These can be Acrobat Capture Assistant stations, devoted entirely to manual steps.

Locating the hub for best performance

Locate the hub on the best machine available—one with a large hard drive and memory, a fast processor or multiple processors, and a fast network interface. Install an Acrobat Capture Cluster Edition station on this hub machine.

In addition, consider enabling the Bind Pages, Mail File, and Store File steps at the hub station, and disabling these steps at other stations in the workgroup. These steps require very little processing time, so they need to be enabled on only one station in a workgroup. Moreover, they work most efficiently when that one station is on the hub machine so that reads and writes between the station and the hub are local ones.

If the performance of a hub machine set up this way is unsatisfactory and you're not in a position to make hardware improvements, consider disabling other steps at the hub station, especially steps that require substantial processing time, such as Capture Image, Export to PDF, and Export to HTML.

Another way to improve hub performance is to have multiple workgroups with multiple hubs. See "Setting up multiple workgroups" on page 59.

Setting up multiple workgroups

You can have as many workgroups as you have Acrobat Capture Cluster Edition and Acrobat Capture Personal Edition stations. The only limit is that Personal Edition stations must be one-station workgroups. Also, with Cluster Edition workgroups, the boundaries between workgroups can easily be altered at need. Any station can join or leave any workgroup with one command.

There are two advantages to having multiple smaller workgroup clusters rather than a single larger one:

- Each cluster can contain a station whose hub is local, thus improving read/write performance on steps such as Store File.
- Workgroup clusters can be configured for high-priority or low-priority workflows, with the high-priority clusters having more processing power. High- and low-priority workflows can run simultaneously in the separate workgroups, with the priority levels of workgroups and workflows matched up.

One disadvantage of separate workgroups is that loads are balanced automatically only within individual workgroups, not among separate ones. One workgroup could have a large backlog of files to process while other workgroups could be sitting idle. Another disadvantage is the added administrative overhead. It may make more sense to maintain a single workgroup and simply stop low-priority workflows when high-priority workflows are running.

Tuning automatic steps

The throughput of workflows with compute-intensive automatic steps such as Capture Image is largely a function of the throughput of these steps. These steps can be tuned to increase their throughput.

The throughput of an automatic step depends on two factors: the number of instances of the step running simultaneously and the processing power available to each instance.

Number of instances of the step Assuming that the step is enabled on all Acrobat Capture Cluster Editions installed in the workgroup, the number of simultaneous instances of the step depends on several factors:

- The number of stations in the workgroup
- Whether the step will run multiple instances on a multiprocessor station (Capture Image, Export to HTML, Export to PDF, Export to TXT, and Store File will run multiple instances.)
- The number of processors licensed for Cluster Edition use at each station (Steps that will run multiple instances at a station can run no more instances than there are licensed processors at the station.)
- The number of instances of the step needed to run to meet workflow demand
- The amount of virtual memory available to each instance of the step

Acrobat Capture will automatically run as many potential instances of the step as are needed if the virtual memory to run them is available. Also, it will shut down instances that are unneeded.

If a potential instance of the step is needed and nevertheless not running, try increasing the virtual memory at its station.

Processing power for each instance The workload is balanced automatically among the processors in a workgroup. If you have a mix of fast and slow processors, however, you can improve the efficiency of automatic load balancing by enabling time-consuming automatic steps such as Capture Image, Export to HTML, and Export to PDF on stations using fast processors and disabling them on slow stations.

Note: *You may also need to disable a third-party step (a step not provided by Adobe) on a machine for which it is not licensed.*

Keeping automatic steps busy

Throughput increases when available steps are busy. Acrobat Capture tries to keep available steps busy by distributing them—putting them to work whenever workflows that use them need the work done.

Because most workflows use a given step only once, this automatic step distribution applies mainly when more than one workflow is active. You can make good use of it by putting multiple workflows to work at the same time and by distributing submission of documents to the different workflows evenly.

Also, you can support automatic step distribution by keeping documents as small as possible. This is because processing can't be distributed among the potential instances of steps that process whole documents. For example, one 100-page document will occupy one instance of the Export to PDF step for as long as it takes the step to finish with the document, even if other potential instances of the step are idle.

Considering limiting and enabling factors

These factors place limits on the configuration of your workgroup or workgroups:

- Each multistation workgroup must have at least one Acrobat Capture Cluster Edition station.
- An Acrobat Capture Personal Edition station cannot join a multistation workgroup.
- A Personal Edition station can use only a single processor, even if it is using a multiprocessor computer.
- In a multistation workgroup, Acrobat Capture Assistant stations cannot perform automatic steps.
- If your workgroup cluster contains over 10 stations, you'll need to locate its hub on a network file server or on a workstation running Windows NT® Server. (NT Workstation permits no more than 10 connections.) Also, if your cluster contains between 5 and 10 stations, you may find that a network file server or NT Server provides better performance.

These factors may enable you to plan smaller, more efficient workgroups than would seem possible at first glance:

- A workgroup member at a single station can control several scanners or submit files from anywhere on the network.
- A Capture Assistant station user can control scanners, submit files, and build workflows as well as perform manual steps.
- Any user of any computer on the network with write access to the workgroup's watched folder, not just users at stations in a workgroup, can place files in the folder.

Calculating hub drive space needs

If you know how much drive space you typically use and you don't have large backlogs of submitted input files, you can calculate the hub drive space you need simply. See "Using the Workgroup panel" on page 51.

Otherwise, your calculations may need to be more elaborate. The hub drive needs enough space to handle three kinds of storage: in-process document storage, input storage space, and output storage space. There are formulas for calculating the size of each of these spaces.

In-process document storage

In-process document storage is working storage for in-process documents. The formula for calculating it is

<system throughput > times < processing delay> times < file size of average input page>

where <system throughput> is the target throughput rate of the workgroup in pages per minute and <processing delay> is the average time in minutes from the time the first page of a document is submitted until the output document is ready.

For example: (50 pages/minute throughput) times (5 minutes processing delay) times (200 KB/page) = 5 MB

Processing delays can be anything from a few seconds for a fully automatic workflow to many minutes for large color documents or workflows with one or more manual steps.

Input storage space

Input storage space is storage space for the maximum number of documents waiting in the input queue. The formula for calculating it is

<maximum input batch> times <system throughput> times <file size of average input page>

where <maximum input batch> is the maximum amount of time in minutes required to process the input batch.

For example: (16 times 60-minute input batch) times (50 pages/minute throughput) times (100 KB/page) = 4,800 MB

In the example, you sometimes scan and submit enough pages in one 8-hour shift to keep your workgroup busy for a full 24-hour day, so 16 hours' worth of documents are sometimes waiting in the input queue at the end of the scanning shift. Maximum input batch is correspondingly smaller if input documents are submitted continuously or at the same rate as system throughput.

Output storage space

Output storage space is temporary storage for finished documents stored in the hub. The formula for calculating it is

<average document storage time> times <system throughput> times < average output document page size>

where <average document storage time> is in minutes (hours times 60).

For example: (4 times 60) times (50 pages/minute throughput) times (15 KB/page) = 180 MB

Output storage space is unnecessary if output documents are not temporarily stored in the hub before being moved to their final destination.

Calculating memory needed for additional processors

One way to eliminate a bottleneck at a step is to add a second processor and run a second instance of the step. When you do this, you need to be sure you have enough memory (physical memory) to keep the new instance performing as well as the original instance.

To calculate the memory required for a processor added to eliminate a bottleneck:

1 Identify the process associated with the bottleneck step. If the step is Capture Image, the process is CapServe.exe. The process for export steps is Export.exe.

2 Submit a typical document to a workflow containing the step.

3 Observe the total system memory usage and the memory allocated to the step's process in Windows NT Task Manager.

4 Add the two memories. The sum is the total physical memory required to maintain the same performance with a second processor or a second pair of processors.

Relocating the hub

The Hub folder can be on any drive accessible to the stations in the workgroup, whether or not the drive is on one of the stations. If the hub you start with is the one installed with your copy of Acrobat Capture Cluster Edition, its default location was in the Acrobat Capture installation folder on your station.

If you decide to move the hub, ideally it should be placed on a fast Windows NT Server drive.

Copying the hub creates a new workgroup. If your workgroup has two Cluster Editions, you might copy the hub and then divide the existing stations between the old workgroup and the new one. You can join more than one workgroup from one station, but only one workgroup at a time.

To move the hub:

1 Stop any workflows that are running.

2 Have all stations disconnect from (leave) the workgroup that uses the hub.

3 Use the Windows file system to move the Hub folder.

4 Rejoin the workgroup whose hub you moved, using the new path to the hub in Acrobat Capture.

5 Run any workflows you stopped.

You can move the hub before the workgroup comes into existence, or at any time afterward. If you move it after other stations have joined, be sure to provide them with the new path to the hub and remind them to rejoin the workgroup at the new path. You may have to share the new Hub folder in order to make it accessible to the other stations.

To copy the hub:

1 Stop any workflows that are running.

2 Have all stations disconnect from (leave) the workgroup that uses the hub.

3 Use the Windows file system to copy the Hub folder.

4 If necessary, share the new Hub folder to make it accessible to other stations.

5 Join the workgroup whose hub is the copy, using the new path to the hub in Acrobat Capture.

6 Cancel all started documents from the new workgroup. (Use Cancel on the Documents panel menu.)

7 Run any workflows you stopped, both in the original workgroup and in the new workgroup.

Identifying workgroup problems

Once you set up your workgroup, your role changes to coach, coordinator, and monitor.

As monitor of your workgroup's resources and performance, you will be troubleshooting workgroup problems as they come up. However, planning and monitoring are a single continuous process. The solutions to problems that turn up while you are monitoring your workgroup are by and large the methods for increasing throughput that you consider when you are planning the workgroup initially. See "Increasing throughput" on page 57.

For monitoring the workgroup, your main resources are your own Workgroup and Station panels and the Station panels at the other stations in the group. For general information on these panels, see "Monitoring steps performed at your station" on page 49 and "Monitoring workgroup performance" on page 50.

Monitoring the Workgroup panel (multistation workgroups)

For monitoring the Workgroup panel, consider the following information.

Hub If you have calculated how much disk space should be available for the hub when Acrobat Capture is idle and how much when a typical number of pages is in the pipeline, you'll have a basis for interpreting the disk space available information provided. See "Calculating hub drive space needs" on page 61.

Station The number of stations, the version of Acrobat Capture installed on each station, and the number of processors on each station provide a highest-level view of the processing resources of your workgroup. You can also observe the memory and disk space available at each station.

As a rule of thumb, the disk space available at a station should be the sum of the maximum space used by all the steps enabled at the station times the number of processors used by the station machine. (The space used by a step is usually two to three times the size of the input file it is processing.)

For information on multiprocessor memory requirements, see "Calculating memory needed for additional processors" on page 62.

Workflow The workflow list includes all workflows used by the workgroup, not just workflows created, run, or used at your station. The count of documents submitted to each workflow provides a good summary of the current backlog of work. You can use it to help you make decisions about temporarily stopping low-priority workflows.

Automatic Step and Manual Step The list includes all steps used by the workgroup, not just steps being performed at your station. The relative number of files waiting for each step is a good indicator of bottlenecks in the system. You can use this count to help you tune steps, for example, by disabling all automatic steps but the overloaded one at stations with fast processors.

Monitoring Station panels

The Station panel at any station provides a lively animated view of the automatic steps being performed at that station. (Multiprocessor stations will show multiple instances of some automatic steps.) The animation identifies the file the step is currently processing. If the step is composed of several phases, it also identifies the current phase.

If your workgroup has several stations in the same room, all visible from one point or during a quick circuit of the room, you can use these animations to monitor all the stations in the group. Just be sure that all the Station panels are in front on all the stations.

When the load on a Capture Image step, a Store File step, or one of the export steps is heavy, a multiprocessor station will launch multiple instances the step. (The number of instances cannot exceed the number of processors used by the station.) If the number is lower than usual, one of its processors may be low on available memory.

Chapter 8

Zoning Images for Capture

Zone Image steps define zones of images, text, and mixed content on the page. You use the Zone Tool application to perform Zone Image steps.

About Zone Image steps

The Capture Image step identifies regions of the page automatically during its page segmentation phase. It then performs word recognition on regions it identifies as text. When a manual Zone Image step precedes a Capture Image step in a workflow, the step provides definitions of regions—*zone definitions*—that override these automatic identifications. For example, advertisements that might otherwise be identified as text regions are defined as image zones so that they are passed over during the word-recognition phase.

Zone definitions can also be used to make up for limitations in automatic region identification. For example, they can identify keyword regions, which can't be identified automatically. And they can identify regions on highly complex or poorly scanned pages when automatic identification of these regions proves unreliable.

For information on page segmentation, see "About zone types" on page 71.

About the Zone Tool application

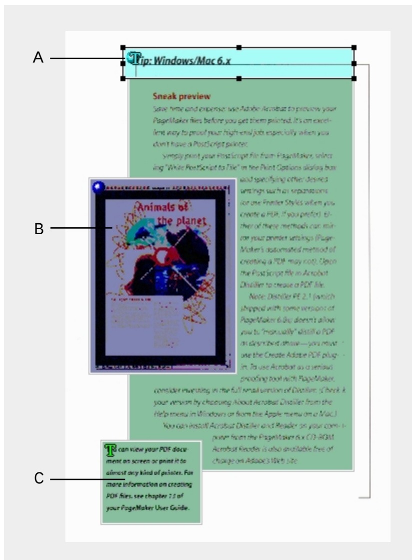
You define zones for a page image by using Zone Tool to outline the regions of the page. Zone definitions are stored in separate zone files.

For information on starting Zone Tool from Acrobat Capture, see "Starting Zone Tool, QuickFix, and Reviewer" on page 3.

Note: *If you will be using Zone Tool extensively and want a printed copy of the instructions for using it, print Chapter 8, "Zoning Images for Capture" (the current chapter) and Chapter 11, "Using metadata, suspects, and dictionaries" in the Acrobat Capture User Guide. The user guide has the same information about Zone Tool as online help, but in a format optimized for printing. The guide is a PDF file (UserGuide.pdf) in the root folder of the Acrobat Capture CD.*

About zone definitions

Zone definitions go into a zone file associated with the image file, and the file is consulted for information as the page is being captured.



A. *Mixed zone* ***B***. *Image zone* ***C.*** *Text zone*

About zone files

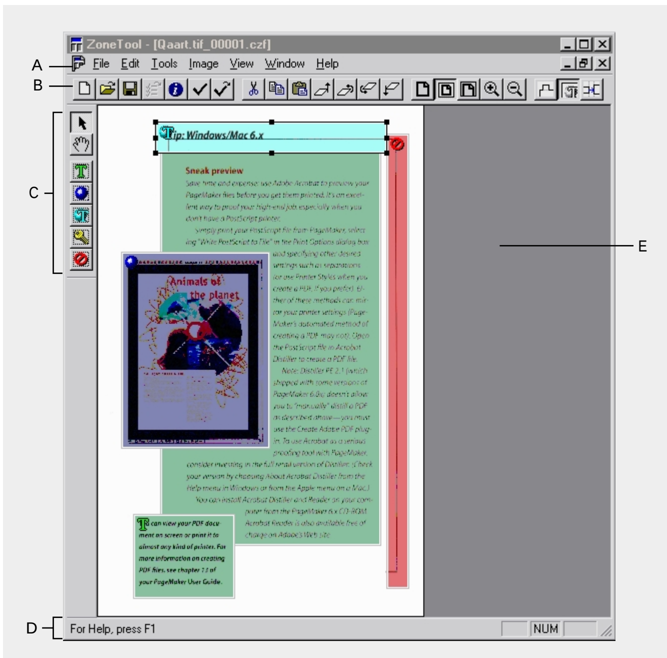
Zone files are created on a per-page basis or developed for use with sets of pages with a similar layout.

• For per-page zone files, a Zone Image step precedes the Capture Image step in the workflow. For the Zone Image step, Zone Tool is used to create a separate zone file for each page submitted to the workflow.

• For a batch of similar pages, you can create a single *default zone file* for use with the Capture Page step in the workflow that will process the pages. In this case, the Zone Image step is omitted from the workflow. For information about creating default zone files, see“Using template and default zone files in workflows” on page 75.

Looking at the Zone Tool work area

The Zone Tool work area has a window for viewing images and defining zones on them, with a menu bar and command bar above the window, a toolbar to the left, and a status bar below.



A. *Menu bar* ***B.*** *Command bar* ***C.*** *Toolbar* ***D.*** *Status bar* ***E.*** *View window*

The command bar, toolbar, and status bar can be hidden as well as shown. The view window has a context menu.

Managing files in a workflow

A document submitted to a workflow containing a Zone Image step arrives at the step as a set of one-page image files. Each page, in sequence, is passed to one of the stations in the workgroup. Using Zone Tool, the operator at a station performs the step by creating zone definitions for the page and committing it back to the workflow. At that point, the file continues to the next step in the workflow.

The operator can also provide document information (metadata) for the file.

At any station running the Zone Image step, any number of files can be waiting for zone definition. The files can be pages from a single document submitted to a single workflow, or pages from separate documents or separate workflows. If page files are waiting at a station when its operator commits a page to the workflow, another page is opened automatically in Zone Tool from a workflow and the process continues.

Acrobat Capture handles the routing of pages to documents and workflows automatically. It deletes the zone file created for a page after the page itself has been fully processed by the workflow.

To open the first available image file from a workflow:

In Zone Tool at any station in the workgroup, do one of the following:

- Click the Open from Workflow button ( ).
- Choose File > Open from Workflow.

If an image file becomes available while Zone Tool is inactive, an “Available!” sound effect plays and the Zone Tool button in the Windows taskbar flashes.

Opening an image file automatically associates a zone file with it.

To activate Zone Tool:

Do one of the following:

- Click the Zone Tool button in the Windows taskbar.
- Click the Zone Image icon in the Station panel of Acrobat Capture.

To provide document information for an image file:

1 Choose File > Document Info.

2 Provide key values as necessary. See “Providing document information (metadata)” on page 119.

To commit an image file to a workflow and open the next available image file:

Do one of the following:

- Click the Commit and Open Next button ( ).
- Choose File > Commit and Open Next.

With this command, you can return an image file you’re finished with to the workflow and at the same time open the next available image file.

Opening an image file automatically associates a zone file with it.

To commit an image file to a workflow:

Click the Commit button ( ), or choose File >Commit.

With this command, you can return the file you’re finished with without opening another one when you’re ready to exit or take a break.

Or, if you aren’t finished with the file, you can save it and finish it later, after reopening it. For instructions, see “Managing files outside a workflow” on page 68.

Managing files outside a workflow

When you use Zone Tool on its own, you employ the ordinary “new,” “open,” and “save” commands to create, open, and save zone files. (The Save Zone File and Save Zone File As commands are disabled when the current file has been opened from a workflow.)

To start Zone Tool outside a workflow:

1 In Acrobat Capture, make sure the Workgroup panel is in front.

2 Do one of the following:

- Double-click Zone Image in the Manual Step section of the panel.
- Select Zone Image, and choose Run Manual Step from the Workgroup panel menu.

You can also start Zone Tool from the Zone Image Properties or Capture Image Properties dialog boxes. For more information, see “Using template and default zone files in workflows” on page 75.

To create a new zone file:

Do one of the following:

- Click the New Zone File button ( ).
- Choose File > New Zone File.

When you create a zone file, a dialog box gives you a chance to display the image whose zones you will define. You can also open the Display Image dialog box later, to provide the image or change it. See “To display a page image for the current zone file:” on page 69.

To open an existing zone file:

Do one of the following:

- Click the Open Zone File button ( ), or choose File > Open Zone File. In the Open dialog box, select the filename, and click Open.
- Choose the document’s filename from the File menu. The menu lists the eight zone files you last opened.
- Double-click the file icon in your file system.

Zone files usually have the extension *.czf*.

When you open a zone file, a dialog box gives you a chance to display the image whose zones you will define. You can also open the Display Image dialog box later, to provide the image or change it. For more information, see “To display a page image for the current zone file:” on page 69.

To display a page image for the current zone file:

1 If the Display Image dialog box is not open, choose File > Display Image.

2 Select an image to display, and click Open. The image can be in TIF, BMP, PCX, JPG, JPEG, or PDF Image format.

To save a zone file:

Do one of the following:

- Click the Save Zone File button ( ), or choose File > Save Zone File.
- Choose File > Save Zone File As, provide a filename, and click Save.

Adjusting page images

When you rotate an image or change its size, its zones rotate or resize with it. Temporarily zooming to a larger image size is useful when you're trying to position or size a zone precisely.

Rotating and resizing change the image you view in Zone Tool. Rotating also changes the image file, but resizing has no effect on it.

To rotate a page image:

Do one of the following:

- To rotate the page image 90˚ clockwise, choose Image > Rotate CW.
- To rotate the page image 90˚ counterclockwise, choose Image > Rotate CCW.
- To rotate the page image 180˚, choose Image > Rotate 180˚.

To change the size of a page image:

Do one of the following:

- To fit the page image to its window, click the Fit Page button ( ), or choose View > Fit Page.
- To fit the page image to the width of its window, click the Fit Width button ( ), or choose View > Fit Width.
- To zoom to the actual size of the page image, click the Actual Size button ( ), or choose View > Actual Size.
- To increase the size of the page image, click the Zoom In button ( ), or choose View > Zoom In.
- To decrease the size of the page image, click the Zoom Out button ( ), or choose View > Zoom Out.

To scroll when the page image is larger than its viewing window:

1 Select the hand tool ( ), or hold down the spacebar.

2 Move the pointer onto the page image.

3 Drag in the direction you want the image to move.

Working with zones

To define a zone for a page image, you draw it on the image. Zones are simple rectangles, but you can complicate their shapes by overlapping or group them into complex *zone clusters*.

Before you draw a zone, you need to identify its type.

About zone types

The zone types are exclude, image, text, keyword, and mixed. Identifying a region of the page as a zone of one particular type overrides automatic identification of the region during a Capture Image step. Defining zones can also improve the accuracy of page segmentation, for example, averting potential confusion where the gap between a block of text and a graphic is narrow enough so that some of the graphic might be interpreted as text.

Note: *Each zone is identified by an icon in the upper left corner, the same icon that appears on the zone tool used to define it. You can turn the display of these icons on and off by clicking the zone icons button ( ) or choosing View > Zone Icons.*

Exclude zones Exclude zones are given the page's background color, for example, the white margins of the scanned pages of a book. They are excluded entirely from page segmentation and word recognition.

Unless you have provided your own template zone file for the Zone Image step in a workflow, the entire zone file is set to Exclude when you first open it. That way you don't have to define margins and other regions of white space as Exclude zones explicitly. (For information on how this background Exclude zone interacts with the zones you define on top of it, see "Layering zones" on page 74.)

Image zones Image zones contain images. In the Capture Image step, they are simply copied to the output file, not processed.

Text zones Text zones contain text. In the Capture Image step, they are converted to characters and words and may be converted to smaller text zones, for example, separate columns.

Keyword zones Keyword zones contain text. In the Capture Image step, keyword zones are treated like text zones but the resulting text is marked as a keyword. The text is included in the document's metadata file, from which any subsequent step can extract keywords. For example, an Export to PDF step can copy any keywords found in the file into the PDF document's Keywords field. For more information, see "Providing document information (metadata)" on page 119.

Mixed Zones Mixed zones contain a mixture of text and images. In the Capture Image step, mixed zones are analyzed and converted to one or more text or image zones according to their content. You would define a zone as mixed if there were no need to perform this analysis and conversion manually.

Defining zones

You define a zone by identifying its type and drawing it. For information on drawing it to the correct size and shape and moving it into the correct position, see "Moving and sizing zones" on page 73.

To define a zone:

1 Select a zone tool from the toolbar or choose a zone tool from the Tools menu. For example, select the Text Zone tool ( ), or choose Tools > Text Zone.

2 Position the pointer at one corner of the zone you want to create, and drag diagonally until the zone is the desired size.

Creating a zone selects it, so you can copy, move, or size it immediately unless you have deselected it first. For information on moving and sizing, see "Moving and sizing zones" on page 73.

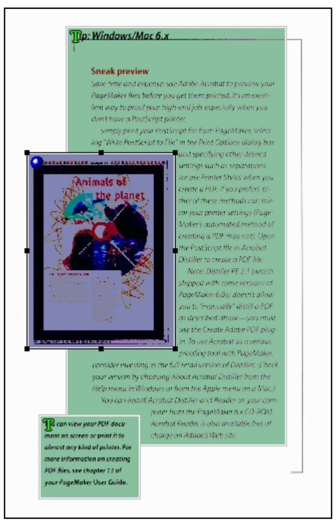
To select a zone:

1 Select the Select tool ( ), or choose Tools > Selection.

2 Do one of the following:

- To select a zone, click the zone.
- To select a zone and display a context menu with relevant commands, right-click the zone.
- To deselect a selected zone, press Shift and click the zone.
- To deselect a selected zone and select the next zone, press Tab.
- To deselect a selected zone and select the previous zone, hold down Shift and press Tab.

The selected zone has size handles and is highlighted.



To select several zones:

Do one of the following:

- To select a zone without deselecting zones already selected, hold down Shift and click the zone.
- To select all of the zones in a zone cluster, hold down Alt and click one of the zones in it.
- To select all zones, choose Edit > Select All.

When two or more zones are selected, the zones are highlighted but have no size handles. For information on zone clusters, see “Working with zone clusters” on page 75.

To define a zone by copying:

1 Select the zone you’re going to copy.

2 Do one of the following:

- Click the Copy button ( ).
- Choose Edit > Copy.

3 Do one of the following:

- Click the Paste button ( ).
- Choose Edit > Paste.

You can create multiple copies by pasting repeatedly.

To define a zone that encompasses the entire page image:

Do one of the following:

- Press Shift while selecting a zone tool from the toolbar.
- Choose a zone type from the Edit > Fill Page With submenu.

To change the type of a zone:

1 Select the zone.

2 Do one of the following:

- Press Ctrl while selecting a zone tool from the toolbar.
- Choose a zone type from the Edit > Change To submenu.

To delete a zone:

1 Select the zone.

2 Choose Edit > Delete or press the Delete key.

Moving and sizing zones

When you move or size a zone, you may want it to overlap or touch another zone. You can move several zones at once, but size only one at a time. For information on overlapping and touching zones, see “Layering zones” on page 74 and “Working with zone clusters” on page 75.

With zone snapping turned on, it’s easier to make zones touch without overlapping.

To turn zone snapping on and off:

Do one of the following:

- Click the snap button ( ).
- Choose Edit > Snap Borders.

When the button is depressed or View > Snap borders is checked, snapping is on. You can also turn snapping on or off temporarily, by holding down the Ctrl key.

To move a zone:

1 Select the zone. If you select several zones, you will move them all at once.

2 Position the pointer anywhere in the zone. The point should change to two crossed double-headed arrows ( ).

3 Drag the zone to its new location.

To nudge a zone while moving it:

1 Select the zone.

2 Press one of the four arrow keys. One press nudges the zone one pixel in the key's direction.

To change a zone's size or proportions:

1 Select the zone.

2 Position the pointer on a handle so that it changes to a single double-headed arrow (⟷).

3 Do one of the following:

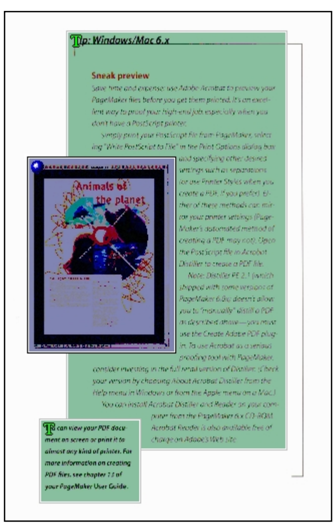
- Drag the handle until the selection is the desired size.
- Hold down the mouse button (as if for dragging) and use one of the four arrow keys to nudge the handle.

As long as the pointer appears as two crossed double-headed arrows ( ), it is positioned for moving the zone, not resizing it.

Layering zones

When you create a new zone overlapping old ones, the new one is placed on top. However, you can alter the layering order in almost any way.

Zones that overlap When zones of different types overlap, the area of overlap is assigned to the zone on top and cut out of the zone or zones under it. You use this feature of zone layering to create "notched" zones and other zone combinations suited to specific pages.



The area of overlap belongs to the image zone on top.

Background zones The one zone whose layering order you can't alter is the zone that forms a background to the entire page image. It remains at the bottom. Unless you use a template zone file for the background, this background is an Exclude zone.

To layer zones:

Do one of the following:

- To bring the selected zone to the top, choose Edit > Bring To Top or click ( ).
- To push the selected zone to the bottom, choose Edit > Push To Bottom or click ( ).
- To move the selected zone up one layer, choose Edit > Move Up or click ( ).
- To move the selected zone down one layer, choose Edit > Move Down, or click ( ).

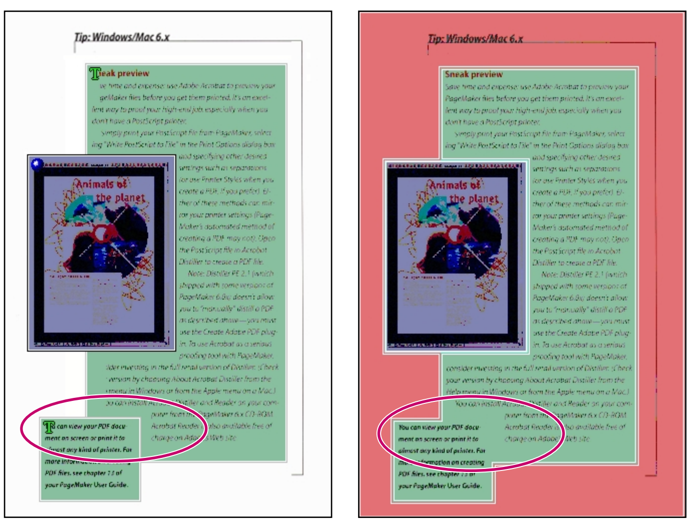
Working with zone clusters

When zones of the same type overlap or touch, they form a zone cluster that is treated as a single object during page capture. You can see the outline of a zone cluster better if you show the zones that form it merged.

To show zones merged:

Do one of the following:

- Press the merge button ( ). Zones are shown merged when the button is depressed.
- Choose View > Zones Merged to activate or deactive zone merging.



Zones in a text zone cluster separate (left) and merged (right)

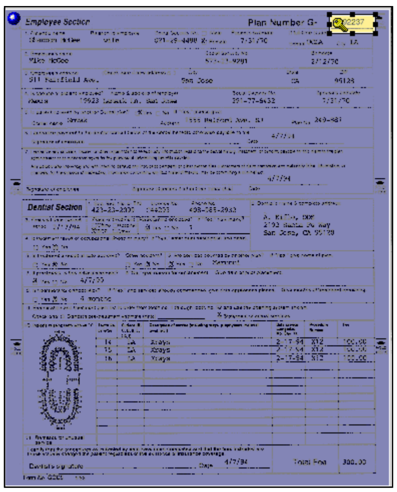
Using template and default zone files in workflows

When the Capture Image step in a workflow needs zone files to be associated with the pages to be captured, you can save labor if the pages are identical or similar in layout. Instead of creating a totally new zone file for each page, you use default zone files with identical pages and template zone files with similar pages.

About default zone files

With a set of pages with the same layout, the same zone file may be applicable to each page. In this case, a default zone file can be used with the Capture Image step, and the Zone Image step omitted from the workflow.

For example, you might create a default zone file for a set of scanned forms that you needed to store as a searchable collection of PDF images. The file would define the preprinted claim number in the upper right corner as a keyword zone and the rest of the form as an image zone.



With the zone file, the body of the form would be ignored during the Capture Page step for each form, and character recognition performed on the claim number only.

Note: *If you have a Capture Page step template or step with a default zone file, you can copy it as is to a workflow with a Zone Image step. The zone file you create in the Zone Image step will override the default zone file.*

About template zone files

With pages that have similar but not identical layouts, each page needs its own zone file but the zone files may have some similarities of zoning. In this case, the workflow can contain a Zone Image step that uses a template zone file containing these similarities. For the step, each page is treated as a separate zoning task in Zone Tool, but the task is quicker and easier because it starts with the template zone file as background rather than the ordinary starting point for a new zone file, a single Exclude zone that has no effect on page segmentation or word recognition.

For example, you might use a template zone file for capturing the pages of a two-column book in which the right column was always text and the contents of the left column varied widely from page to page. The template would define the right column as a text zone so that anyone providing per-page zone files for the book would need to work on the left columns only.

Setting up template and default zone files

You can provide template or default zone files for steps in specific workflows, or for step templates. For information about the difference in effect, see “Configuring Steps” on page 17.

To set up a template zone file with a Zone Image step:

1 Open the Zone Image Properties dialog box. For instructions, see “Changing step configurations” on page 17.

2 Click New opposite Template zone file. This starts Zone Tool with a new zone file.

3 Create the zone file.

4 Save the file in the folder specified (Hub\Configs).

5 Exit Zone Tool.

6 In Zone Image Properties, choose the file you saved as the template zone file.

7 Click OK.

To set up a default zone file with a Capture Image step:

1 Open the Capture Image Properties dialog box. For instructions, see “Changing step configurations” on page 17.

2 Click New opposite Default zone file. This starts Zone Tool with a new zone file.

3 Create the zone file.

4 Save the file in the folder specified (Hub\Configs).

5 In Capture Image Properties, choose the file you saved as the default zone file.

6 Click OK.

For more information about Capture Image properties, see “Configuring Capture Image” on page 19.

To edit a template or default zone file:

1 Do either of the following:

- To edit a template zone file, open the Zone Image Properties dialog box.
- To edit a default zone file, open the Capture Image Properties dialog box.

2 Select a zone file.

3 Click Edit opposite Template zone file or Default zone file. This starts Zone Tool with the selected file open in it.

4 Edit and save the file.

5 In the properties dialog box, Click OK.

Zone Tool shortcuts

Most shortcuts listed here have no equivalent commands. For the shortcut to a command on a menu, see the menu. For a shortcut to a tool, see the tool's tip.

Selecting tools	
Result	**Action**
Select Text Zone tool	1 or T
Select Image Zone tool	2 or I
Select Mixed Zone tool	3 or M
Select Keyword Zone tool	4 or K
Select Exclude Zone tool	5 or E
Select hand tool	Depress spacebar or H

Selecting zones	
Result	**Action**
Deselect all zones	Escape
Toggle selection	Shift-click
Select previous zone	Shift-Tab
Select next zone	Tab

Moving around	
Result	**Action**
Scroll left and right	Ctrl+Page Up, Ctrl+Page Down
Scroll up and down	Page Up, Page Down

Using the mouse	
Result	**Action**
End drag	Escape

Editing zones	
Result	**Action**
Change selected zones to tool type	Ctrl+click zone tool
Fill page with tool type	Shift+click zone tool
Nudge zone	Arrow keys

Using the Clipboard	
Result	**Action**
Copy	Ctrl+C or Ctrl+Insert
Cut	Ctrl+X or Shift+Delete
Paste	Ctrl+V or Shift+Insert

Chapter 9
QuickFixing Word Suspects

You use QuickFix Page steps to check and fix suspects. You use the QuickFix application to perform QuickFix Page steps.

About QuickFix Page steps

Suspect words are potential errors recognized during a Capture Image step that may need to be revised, changed back to the original bitmap, or deleted. Using a QuickFix Page step in a workflow is the quickest and easiest way to check and fix suspects flagged by an earlier Capture Image step. For information, see “Working with suspects” on page 121.

Use a Review Document step rather than QuickFix Page if you need to check for formatting or layout errors as well as suspect words. If you use neither QuickFix Page nor Review Document, you can still substitute the original bitmap for all suspects in an Export to PDF step. For information, see “Configuring Export to PDF” on page 22. For information on the Review Document step and Reviewer, the application used in it, see “Reviewing Captured Text and Graphics” on page 93.

About the QuickFix application

QuickFix simplifies checking and correction of the word suspects in a captured page (ACP) file by displaying the suspects in a table rather than displaying the whole page.

For information on starting Reviewer from Acrobat Capture, see “Starting Zone Tool, QuickFix, and Reviewer” on page 3.

Note: *If you will be using QuickFix extensively and want a printed copy of the instructions for using it, print Chapter 9, “QuickFixing Word Suspects” (current chapter) and Chapter 11, “Using metadata, suspects, and dictionaries” in the Acrobat Capture User Guide. The user guide has the same information about QuickFix as online help, but in a format optimized for printing. The guide is a PDF file (UserGuide.pdf) in the root folder of the Acrobat Capture CD.*

About suspects in QuickFix

For each suspect word, the QuickFix table includes the source image (original bitmap region), the word itself, and the reason the word is suspect. The reasons listed depend on QuickFix Page property settings made in the workflow configuration. By default, they are low confidence, absence from the dictionary, uncertain font, numeric characters, and alphanumeric characters. For information, see “Working with suspects” on page 121.

You check a suspect word by comparing it to the bitmap and considering the reason it is suspect. You fix the page by skimming the rows of the table, fixing errors and accepting words that are correct. When you accept or delete a suspect, or change it back to its bitmap image, its row is removed from the table.

Suspect rows are ordered by the location of the word image on the page, but you can also sort them alphabetically by suspect word, by confidence, or by reason. These sorts are useful for grouping similar suspects. Grouping streamlines the fixing process.

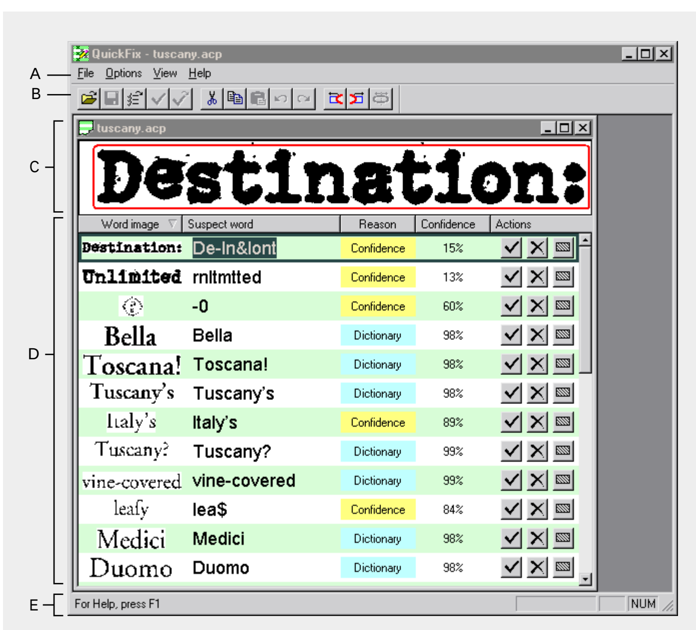
About QuickFix’s input and output

QuickFix works directly with the single-page ACP files produced in the Capture Image step. If the original pages were part of a multipage document, this means the task of checking word suspects in the document can be divided among more stations in the workgroup and thus performed more quickly. For more information on workgroup planning, see “Managing Workgroups” on page 55.

Whether or not the pages captured in a Capture Image step are processed at the same station, QuickFix treats any set of them that are part of the same document as a unit. For example, accepting a Dictionary suspect at any station automatically accepts all instances of the same word that are Dictionary suspects, and not just on the page where it appears but on all further pages in the document not yet processed in QuickFix.

Looking at the QuickFix work area

The QuickFix work area has a window for viewing captured pages and fixing suspects on them, with a menu bar and command bar above the window and a status bar below. The window has a lower pane that lists suspect words and an upper pane that displays a bitmap image of the page region containing the suspect selected from the list.



A. *Menu bar* ***B.*** *Command bar* ***C.*** *Upper pane* ***D.*** *Lower pane* ***E.*** *Status bar*

The command bar and status bar can be hidden as well as shown. The suspect list in the lower pane has a context menu.

About the suspect list

The suspect list is in the form of a table, with each suspect and and related information making up one row.

In addition to the suspect word, the row contains a bitmap of the word (word image); the reason why the word is a suspect; the word's recognition confidence, expressed as a percentage (100% would be total confidence); and three of the actions you can perform on the suspect word: accept it ( ), delete it ( ), and change it to an image ( ).

Adjusting the work area

You can show or hide the menu bar and command bar and change the width of columns.

To change the width of a column:

Drag the border between the two columns.

You can change the width of the Word Image and Suspect Word columns only.

Managing files in a workflow

A document submitted to a workflow containing a QuickFix Page step arrives at the step as a set of captured page (ACP) files. Each captured page, in sequence, is passed to one of the stations in the workgroup. Using QuickFix, the operator at a station performs the step by checking and correcting suspect words on the page and committing the file back to the workflow. At that point, the file continues to the next step in the workflow.

At any station running the QuickFix Page step, any number of ACP files can be waiting for checking and correcting. The files can be pages from a single document submitted to a single workflow, or pages from separate documents or separate workflows. If files are waiting at a station when its operator commits a file to the workflow, another file is opened automatically in QuickFix from a workflow and the process continues. Acrobat Capture handles the routing of pages to documents and workflows automatically.

To open the first available ACP file from a workflow:

In QuickFix at any station in the workgroup, do one of the following:

- Click the Open from Workflow button ( ).
- Choose File > Open from Workflow.

If an ACP file becomes available while QuickFix is inactive, an "Available!" sound effect plays and the QuickFix button in the Windows taskbar flashes.

To activate QuickFix:

Do one of the following:

- Click the QuickFix button in the Windows taskbar.

• Click the QuickFix Page icon in the Station panel of Acrobat Capture.

To commit an ACP file to a workflow and open the next available ACP file:

Do one of the following:

• Click the Commit and Open Next button ( ).

• Choose File > Commit and Open Next.

With this command, you can return the file you're finished with and open the next available ACP file.

To commit an ACP file to a workflow:

Do one of the following:

• Click the Commit button ( ).

• Choose File > Commit.

With this command, you can return the file you're finished with without opening another one when you're ready to exit or take a break.

To close and relinquish a file:

1 Choose File > Close.

2 If you are given the choice of saving or not saving changes, click Yes or No.

If you aren't finished with the file, you can close and relinquish it so that you (or another workgroup member) can reopen it from the workflow and finish it later.

Sorting suspect lists

You can sort suspects by location on the page, in alphabetical order, by the reason they are suspect, or by the confidence (expressed as a percentage) that they have been correctly recognized.

An alphabetical sort is useful when the same word appears several times on the page with the same error.

Sorting by confidence is a good way to isolate suspects likely to require correction.

Sorting by reason is useful when one reason is particularly important. For example, you might want to group all Numeric suspects in an important financial document so that you could focus on the numbers. In this case, you would probably consider the recognition confidence irrelevant.

For information on correcting and accepting suspects in groups, see "Performing repeated and global edits" on page 89.

To change the sort order of suspect words:

Do one of the following:

• To sort by the location of words on the page, click the button at the top of the Word image column, or choose Options > Sort By Location.

- To sort alphabetically, click the button at the top of the Suspect word column, or choose Options > Sort Alphabetically.
- To sort by the reasons the words listed are suspect, click the button at the top of the Reason column, or choose Options > Sort By Reason. Suspects that have the same reason are sorted in increasing confidence order.
- To sort by confidence, click the button at the top of the Confidence column, or choose Options > Sort By Confidence. Suspects that have the same confidence percentage are sorted by location on the page.

The row selected at the time you sort remains selected and visible, even if that requires that the suspect list scrolls up or down to display it.

Editing suspect words

Editing options include accepting, deleting, and revising the suspect word, and changing the word to an image.

About suspect edits

You can edit any suspect word in the following ways.

Accept the suspect word. Accepting a suspect word has three effects:

- Removes the word from the suspect list.
- Makes the word non-suspect for any steps later in the workflow, even if it qualifies as a suspect by their criteria.
- If the word is a Dictionary suspect, automatically accepts any other instances of the same Dictionary suspect in the document. For more information on dictionaries, see “Adding QuickFix suspects to dictionaries” on page 88.

Delete the suspect word What you delete is usually not a word at all, but stray pixels or part of a graphic incorrectly recognized as a word.

Deleting removes the suspect and its row from the suspect list and discards the suspect. If the page is converted to PDF Formatted Text & Graphics, HTML, or TXT later in the workflow, the original bitmap image of the suspect will not be displayed either. (With TXT conversions, but not PDF Formatted Text & Graphics or HTML conversions, the space it occupied will be closed up.)

If the page is converted to PDF Searchable Text, the original bitmap of the suspect, as part of the bitmap of the entire page, is displayed as is.

Change the suspect word to an image You would make this change if the image was a signature, or if it was an in-line graphic mistaken for a word.

Changing to image, like deleting, removes the suspect and its row from the suspect list and discards the suspect. But if the page will be converted to PDF Formatted Text & Graphics or HTML later in the workflow, it also replaces the suspect on the page with its bitmap image.

For pages that will be converted to TXT or PDF Searchable Image, changing the suspect to an image has the same effect as deleting it.

Revise the word You can make any of the following revisions to the suspect word:

- Replace the word with a word you enter.
- Insert characters in the word.
- Delete selected characters in the word, or replace them with characters you enter.
- Insert a special character in place of or within a word. Special characters are more likely to be misinterpreted in the Capture Image step than regular characters.
- Merge the word with the previous or next word. This is useful when a single word has been incorrectly recognized as two words.
- Split the word into two words. When the space between two words not listed in the dictionary is small, they're sometimes run together as a single word.

None of these revisions are accepted automatically. You need to accept a revised suspect as a separate step.

Performing suspect edits

Most suspect edits can be performed in two steps, selecting the word and editing it. With the more common edits, the two steps can be accomplished with one click.

To select a suspect word:

Do any of the following:

- To select any suspect word, click in its row.
- To select the next word on the list, press the Down Arrow key or the Tab key.
- To select the previous word on the list, press the Up Arrow key or hold down Shift and press Tab.
- To move word selection to the previous or next screen of items, press Page Up or Page Dn.
- To move word selection to the first or last item on the list, hold down Ctrl and press Home or End.

To accept a suspect word:

Do one of the following:

- Click the accept button ( ) in its row.
- Select its row and choose Edit > Accept Word.
- Select its row and press Enter.

To delete a suspect word:

Do one of the following:

- Click the delete button ( ) in its row.
- Select its row and choose Edit > Delete Word.

To change a suspect word to its bitmap image:

Do one of the following:

- Click the image button ( ) in its row.
- Select its row and choose Edit > Change to Image.

To replace a suspect word:

1 Select the row containing the word. (You don't have to select the word itself.)

2 Do one of the following:

- Type or paste the new word.
- Use a special character.

To revise a suspect word:

1 Select the row containing the word.

2 Do one of the following:

- To replace characters in a suspect word, select the characters.
- To insert characters in a suspect word, click where you want the insertion.

3 Do one of the following:

- Type or paste the new characters.
- Use a special character.

To use a special character:

1 Do one of the following:

- Choose a special character such as Em dash (—) or Copyright (©) from the Insert menu.
- Choose Other from the Insert menu.

2 If you chose Other, do one of the following:

- Double-click a character in the Symbol list.
- Select a character from the Symbol list and click Insert.

Rather than selecting curly quotation marks from the Symbol list, you can simply type them.

To set up curly quotation mark typing:

Choose Options > Smart Quotes to activate Smart Quotes.

When Smart Quotes is activated (checked), single and double quotation marks you type are curly.

Curly quotation marks are open at the beginning of a word and closed later in the word—ordinarily at the end of the word.

To merge a suspect word with the next or previous word:

1 Select the row containing the suspect word.

2 Do one of the following:

- Click the Merge Next button ( ), or Choose Edit > Merge with Next.

• Click the Merge Previous button ( ), or Choose Edit > Merge with Previous.

To split a suspect word into two words:

1 Select the row containing the suspect word.

2 In the row, click the suspect word where you want it to split.

3 Do one of the following:

• Click the Split Word button ( ), or choose Edit > Split at Cursor.

• Enter a space character (press the spacebar).

If you click Split Word or choose Edit > Split at Cursor, the word is split in two and placed in two separate rows. If you enter a space, the word is split only after you accept it or save the file.

Undoing or redoing an edit

You can undo or redo the edits you have performed, and you can redo any edit you have undone.

To undo edits:

1 Click the Undo button ( ), or choose Edit > Undo. The effect is to undo your most recent edit.

2 If necessary, repeat the step. The effect is to undo the next most recent edit, and so on. If necessary, you can continue undoing until you reach the original (last saved) version of the file you are editing.

To redo edits:

1 Click the Redo button ( ), or choose Edit > Redo. The effect is to redo the edit you have recently undone.

2 If necessary, repeat the step. The effect is to redo the next most recent “undo,” and so on.

Adding QuickFix suspects to dictionaries

The Capture Image step that produces the pages you work with in QuickFix uses at least one standard language dictionary for OCR purposes. It can also use user dictionaries for specialized word lists. QuickFix identifies a word as a Dictionary suspect only if it doesn’t appear in any of these standard or user dictionaries.

You can edit these user dictionaries while you are working in QuickFix. Alternatively, QuickFix provides a quick way of adding suspect words to user dictionaries. As well as adding to the dictionary for future use, the quick method has the immediate effect of accepting the suspect word and all other instances of the suspect in the document. For information, see “Working with user dictionaries” on page 122.

Note: *If the words you want to add are Dictionary suspects that will be useful only for the document you are processing, not for future documents, don’t bother to add them to a user dictionary. Just accepting a Dictionary suspect will accept all other instances of that suspect in the document.*

Performing repeated and global edits

Accepting a Dictionary suspect automatically accepts all instances of that suspect. However, accepting doesn't have this sort of global effect with other kinds of suspect words. For them, you make repeated replacements of suspects by duplicating or pasting. The word you duplicate or paste is ordinarily itself an edit, one that you want to repeat because it corrects a suspect word that is repeated several times on the page.

Duplicating repeats the edit on the next item in the list. When used after an alphabetical sort, it's the best method for pages where the same word appears several times with the same error. Pasting repeats the edit on any suspect you select.

To duplicate an edit with the next word or words on the list:

1 Select the row containing the word you want to duplicate.

2 If you need to edit the word before you duplicate it, perform the edit but do not accept it. (Accepting it at this point would remove it from the suspect list, making it inaccessible.)

3 Choose Edit > Duplicate To Next to duplicate the word on the next item in the list. Repeat the step for as many consecutive ("next") items as you want to edit in the same way. Accept the duplicates only after you have finished duplicating.

To paste an edit in another suspect:

1 Select the row containing the word you want to paste.

2 If you need to edit the word before you paste it, perform the edit but do not accept it. (Accepting it at this point would remove it from the suspect list, making it inaccessible.)

3 Click the Copy button ( ), or choose Edit > Copy. (At this point, you can accept the word you copied.)

4 Select the suspect you want to edit.

5 Click the Paste button ( ), or choose Edit > Paste.

You can accept all suspects in the list at once. With pages on which most of the suspect words are acceptable, it may be faster to look through the list for deletions and changes to make, make them, and then accept all the remaining words at one time.

To accept all the suspect words on the page:

Choose Edit > Accept All.

Developing editing strategies

The approach you take to editing a captured page depends on the kind of suspects it contains and your purpose for editing it. For example, you will be more likely to proof Numeric suspects carefully if a scan prior to the Capture Image step was a poor one and the numbers are important ones.

To fix a page containing clean (well scanned) text:

1 Sort by Reason, so that Dictionary and Numeric suspects are grouped.

2 Review the group of Dictionary suspects, and accept correct words. Because accepting a Dictionary suspect automatically accepts all other instances of the same suspect, this may shorten the suspect list quickly.

3 Review the group of Numeric suspects briefly, and accept them. The page is well scanned, so these suspects probably don't need correcting.

4 Sort by Suspect word (alphabetically), and review repeated suspects, copying and pasting, or duplicating, when you can.

5 Sort by Confidence to bring low-confidence suspects to the top of the list.

6 Review any suspects remaining, sorting by Word image (location) if position on the page is useful information in the review.

To fix a page for conversion to TXT or PDF Searchable Image quickly:

1 Sort by Reason.

2 Accept all Dictionary suspects, first correcting when necessary.

3 Correct other suspects as necessary, but do not accept them.

4 Commit the page to the workflow.

Accepting Dictionary suspects can save time, because it automatically removes other instances of the same suspects from the suspect list. Otherwise, accepting has no effect on the display in TXT or PDF Searchable Image and takes up time.

To fix pages as quickly as possible:

Do any of the following:

• Use only the accept ( ), delete ( ), and image ( ) buttons for editing. (Don't type corrections, insert special characters, or merge or split words, all of which are relatively time-consuming activities.)

• Don't edit inessential types of suspects. For example, leave all numbers as suspects when it is unlikely that there will be searches for numbers in the document. If you are exporting to PDF or HTML, the original images will be displayed, so any recognition errors will be immaterial.

QuickFix shortcuts

Most shortcuts listed here have no equivalent commands. For the shortcut to a command on a menu, see the menu.

Editing	
Result	**Action**
Undo	Alt+Backspace or Ctrl+Z
Redo	Alt+Shift+Backspace or Ctrl+Y
Accept word	Enter or Ctrl+Tab

Editing	
Result	**Action**
Copy	Ctrl+C or Ctrl+Insert
Cut	Ctrl+X or Shift+Delete
Paste	Ctrl+V or Shift+Insert

Chapter 10

Reviewing Captured Text and Graphics

Review Document steps check and correct formatting or layout errors as well as suspect words. You use the Reviewer application to perform Review Document steps.

About Review Document steps

Use a Review Document step if you need to check for formatting or layout errors as well as suspect words. (Suspect words—*suspects*—are potential errors flagged during a Capture Image step that may need to be revised, changed back to the original bitmap, or deleted. For information, see “Working with suspects” on page 121.

Use a QuickFix Page step if you need to check for suspect words only. If you use neither QuickFix Page nor Review Document, you can still substitute the original bitmap for all suspects in an Export to PDF step. For information, see “Configuring Export to PDF” on page 22. For information on the QuickFix Page step and QuickFix, the application used in it, see “QuickFixing Word Suspects” on page 81.

About the Reviewer application

Reviewer works with files that have been processed by a Capture Image step and a Bind Pages step. Capture Image performs OCR on the individual pages of a document, and Bind Pages reassembles the pages into an ACD file. You view the ACD file in Reviewer, a page at a time. The page you see is exactly what the Capture Image step has outputted, with all recognition errors visible and all suspects highlighted.

Suspect highlighting is color-coded, with a different color for each reason a word might be suspect. The colors (reasons) shown depend on Review Document property settings made in Acrobat Capture. By default, they are the colors for confidence, dictionary, font, numeric, and mixed numeric. For information, see “Working with suspects” on page 121. If the suspect property settings are good ones, all errors are highlighted.

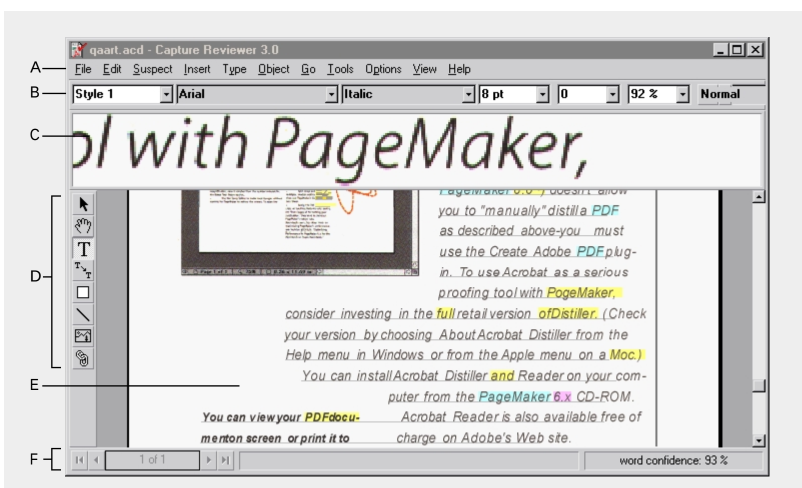
You use Reviewer to correct errors in layout and image recognition as well as word recognition errors. You can also use it for certain enhancements to documents—updated illustrations, hidden keywords for images, new lines of text, numerical ordering of text lines to improve accessibility for visually impaired readers, and so on.

For information on starting Reviewer from Acrobat Capture, see “Starting Zone Tool, QuickFix, and Reviewer” on page 3.

Note: *If you will be using Reviewer extensively and want a printed copy of the instructions for using it, print Chapter 10, “Reviewing Captured Text and Graphics” (current chapter) and Chapter 11, “Using metadata, suspects, and dictionaries” in the Acrobat Capture User Guide. The user guide has the same information about Reviewer as online help, but in a format optimized for printing. The guide is a PDF file (UserGuide.pdf) in the root folder of the Acrobat Capture CD.*

Looking at the Reviewer work area

The Reviewer work area is a window for viewing and reviewing a captured document, with a menu bar, format bar, toolbar, and status bar around the outside of the view. You view the document in two panes, a document pane showing a page of the document and a context pane showing the original bitmap image of a word in its context—together with the words before and after it. The word shown is the one the cursor is located on.



A. *Menu bar* ***B.*** *Format bar* ***C.*** *Context pane* ***D.*** *Toolbar* ***E.*** *Document pane* ***F.*** *Status bar*

The format bar, toolbar, and context pane can be hidden as well as shown. The document pane has a context menu.

To make tool selections persist after the tool is used:

Choose Tools > Persistent Selection to activate or deactivate persistent selection.

Ordinarily, the selection tool ( ) is selected until you select another tool and reselected after you have used the other tool. Only the selection tool remains selected after use. With persistent selection active, all tools (not just the selection tool) remain selected after you use them.

Managing files in a workflow

A document submitted to a workflow containing a Review Document step arrives at the step as a captured document (ACD) file. The file is passed to one of the stations in the workgroup. Using Reviewer, the operator at a station performs the step by reviewing the document and committing the ACD file back to the workflow. At that point, the file continues to the next step in the workflow.

The operator can also can provide document information (metadata) for the file.

At any station running the Review Document step, any number of ACD files can be waiting for review. The files can be separate documents submitted to a single workflow, or documents submitted to separate workflows. If ACD files are waiting at a station when its operator commits a file to the workflow, another file is opened automatically in Reviewer from a workflow and the process continues. Acrobat Capture handles the routing of documents to workflows automatically.

To open the first available ACD file from a workflow:

In Reviewer at any station in the workgroup, choose File > Open from Workflow.

If an ACD file becomes available while Reviewer is inactive, an “Available!” sound effect plays and the Reviewer button in the Windows taskbar flashes.

To activate Reviewer:

Do one of the following:

- Click the Reviewer button in the Windows taskbar.
- Click the Review Document icon in the Station panel of Acrobat Capture.

To provide document information for an ACD file:

1 Choose File > Document Info.

2 Provide key values as necessary. For information, see “Providing document information (metadata)” on page 119.

To commit an ACD file to a workflow and open the next available ACD file:

Choose File > Commit and Open Next.

With this command, you can return a reviewed ACD file to the workflow and open the next available ACD file.

To commit an ACD file to a workflow:

Choose File > Commit.

With this command, you can return the reviewed ACD file without opening another one.

To close and relinquish a file:

1 Choose File > Close.

2 If you are given the choice of saving or not saving changes, click Yes or No. (You can confirm changes if Options > Confirm Changes on Close is checked.)

If you aren't finished with the file, you can close and relinquish it so that you (or another workgroup member) can reopen it from the workflow and finish it later.

Viewing documents

When you open a document, you view it in both the context pane and the document pane.

Viewing a document in the document pane

You can set the default zoom—the magnification at which Reviewer displays pages when it opens a document in the document pane. Whatever the default, you can also zoom in or out while you are viewing the document.

One possible viewing adjustment is to "autozoom"—zoom in on a particular word to view it at a point size you have prespecified for easy reading. Another is to choose to display, or not display, large bitmap images. Not displaying them speeds page display.

To change the zoom while viewing a document:

Do any of the following:

• To zoom in, choose View > Zoom In. The magnification level is increased so you can see more detail.

• To zoom out, choose View > Zoom Out. The magnification level is decreased.

• To display the page at 100% magnification, choose View > Actual Size.

• To display the entire page, choose View > Fit in Window.

• To match the width of the page to the width of the Reviewer window, choose View > Fit Width.

• To match the width of the page content to the width of the Reviewer window, choose View > Fit Visible. *Page content* is the block of text, art, or text and art inside the margins.

To view another part of a page larger than the document pane:

1 Select the hand tool ().

2 Drag the page away from the part you want to view. For example, if you want to view a lower part of the page magnified to a size larger than the document pane, drag the page upward.

Moving a page with the hand tool is like moving a piece of paper on a desk with your hand.

To zoom in on a particular suspect word:

1 Click the word.

2 Choose View > AutoZoom.

To change the zoom at which documents are opened:

Do one of the following:

- Choose Options > Default Zoom > Fit in Window.
- Choose Options > Default Zoom > Actual Size.
- Choose Options > Default Zoom > Fit Width.

To calibrate autozooming:

1 Choose Options > Calibrate AutoZoom.

2 Specify the preferred text size in points.

3 Click OK.

To display large bitmap images:

Choose Options > Display Large Images to activate or deactivate display of large images.

When Display Large Images is inactive (unchecked), a gray box is shown in place of the image.

To autozoom automatically, on the next suspect selected:

Choose Options > AutoZoom on Next Suspect to activate or deactivate automatic autozoom.

Viewing a document in the context pane

You can zoom out or scroll to view a larger context.

To expand the context:

1 Move the cursor to the context pane.

2 Press and hold the mouse button. The pane zooms out so that more text is visible.

3 To scroll the text in the pane, drag upward or downward.

4 Release the button to zoom back in to the original context.

Navigating a document

Reviewer provides a number of ways to go from one page to another, all described in this section. In addition, simply tabbing from the last suspect on a page displays the following page.

To page through a document:

Do one of the following:

- Choose Go > Next Page or Go > Previous Page.

- Choose Go > First Page or Go > Last Page.
- Choose Go > Back or Go > Forward.

To go to a specific page:

Do one of the following:

- Click the page number button in the status bar at the bottom of the window, type the page number, and press Enter.
- Choose Go > Go To Page, type the page number, and click OK.

Checking and correcting suspects

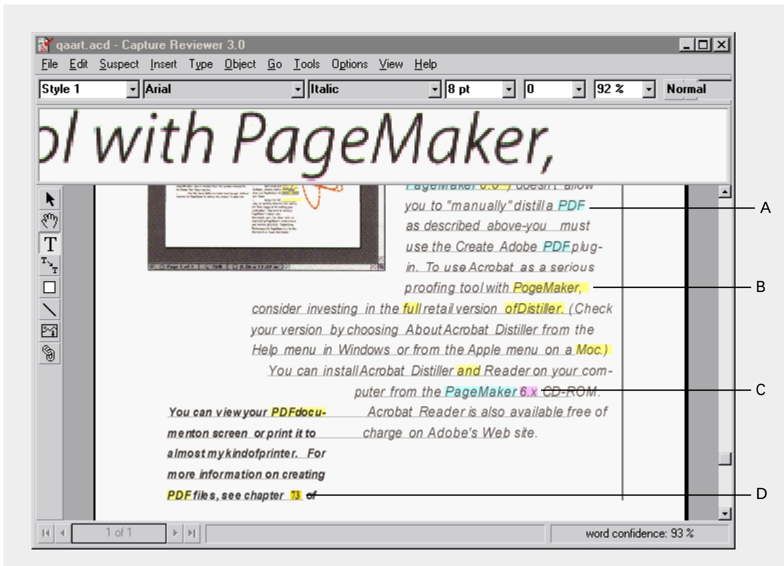
When you open an ACD document with Reviewer and select the type tool, suspect words are highlighted. When you select a particular suspect, the status bar provides further information about it.

If you accept a suspect word, it loses its suspect status and will be displayed as regular text when the ACD document is exported to PDF or HTML. If you leave it suspect, the original bitmap image will be displayed instead. This is the case whether or not you edit the word first (that is, before you accept it or leave it suspect).

About suspect highlights

Words rated as suspects because they fall below the word confidence threshold are highlighted with a yellow background.

Reviewer can also highlight words that are not found in the dictionary, words with unidentifiable fonts, words that contain a mix of alphabetic and numeric terms, and all numeric terms. These kinds of suspects are highlighted with blue, green, magenta, and orange backgrounds, respectively.

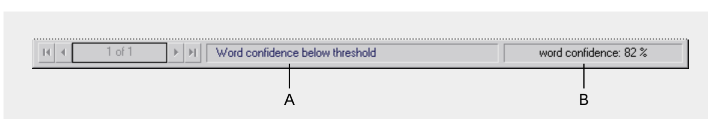


A. *Dictionary (blue)* ***B****. Confidence (yellow)* ***C****. Alphanumeric (magenta)* ***D.*** *Numeric (orange)*

The suspect settings determine which kinds of suspects are highlighted. The word confidence highlight threshold determines how many recognition suspects are highlighted.

About suspect status

The Word Confidence area in the status bar shows the word confidence level of the selected suspect. The Hint area shows the type of suspect word—word confidence below threshold or one of the other types.



A. *Hint area* ***B.*** *Word Confidence area*

Performing checks and corrections

When you check a suspect, you can accept it as is, correct it and then accept it, or leave it as asuspect.

To check and correct suspects:

1 Make sure the type tool (T) is selected.

2 Press the Tab key to jump to the first suspect. The original bitmap image of the suspect appears in the upper pane. (If there is no upper pane and you want to see the original bitmap, choose View > Context Pane.)

3 If necessary, edit the suspect. For information, see “Editing text” on page 100.

4 Do one of the following:

• To accept the suspect word as correctly recognized and jump to the next suspect, choose Suspect > Accept & Find Next.

• To accept a Dictionary suspect and all other occurrences of the same word in the current document, choose Suspect > Add to Dictionary. For information about Add to Dictionary, see “Adding Reviewer suspects to dictionaries” on page 104.

• To jump to the next suspect without accepting this one, press Tab or choose Suspect > Find Next.

5 Continue tabbing to suspects, either accepting them as is, correcting them and then accepting them, or leaving them as suspects as you proceed through the document.

Editing text

You can edit by inserting new text or by changing existing text—deleting it, replacing it, reformatting it, or converting it. Ordinarily you will edit only individual suspect words. However, you can edit any word in the document, whether or not it is suspect, and any line of text. Also, you can create new text lines. For information about text lines, see “Creating and editing text lines” on page 109.

In most cases, you have to select existing text before you can edit it.

Selecting text

You use the type tool (T), not the selection tool, for text selection.

To select text:

1 Select the type tool.

2 Do one of the following:

• To select any text, drag through it.

• To select a word, double-click it. (For changing character format, placing the cursor anywhere in a word selects the word.)

• To select a line, triple-click it.

• To select the character to the left or right of the cursor, press Shift+Left Arrow or Shift+Right Arrow.

• To select or deselect all the text on the page, choose Select > Select All or Select > Deselect All.

Inserting or replacing text

If you want to replace text, you need to select it before you type the replacement.

To insert or replace text:

1 Do one of the following:

• To insert text, use the type tool (T) to position the I-beam cursor at the location where you want to insert text, and click to position the insertion-point cursor.

• To replace text, use the type tool to select the characters to replace. Select no more than a single line at a time. You can't replace two or more lines at once.

2 Type or paste the text, or insert a special character. For instructions, see "Using symbol characters as insertions or replacements" on page 103.

If the same text is mistakenly recognized several times in a document, you can find and replace all instances of it at once.

To find and replace text:

1 Choose Edit > Find/Change.

2 Choose one, neither, or both of the following options:

• **Match whole word only** does not find subsets of characters in words. For example, if you enter *now*, it is not found in the word *nowhere*.

• **Match case** finds text only when the characters match exactly. For example, the words *Now* and *now* are not an exact match.

3 Type the text to find in the Find What text box, and click Find Next.

4 Type the replacement text in the Replace With text box, and click Replace.

5 Continue clicking Replace until all appropriate words have been replaced.

6 Click Close.

Deleting text

You can delete a selection of characters or delete characters one by one. You can also undo the deletions.

To delete characters:

1 Use the type tool (T) to select the characters to delete.

2 Press Backspace or Del. (You may have to press Num Lock to use the Del key.)

For deleting characters one by one, you don't have to select them first.

To delete characters one by one:

1 Make sure the type tool (T) is selected.

2 Position the insertion-point to the left or right of the first character you want to delete.

3 Do one of the following:

• To delete the character to the left of the insertion point, press Backspace.

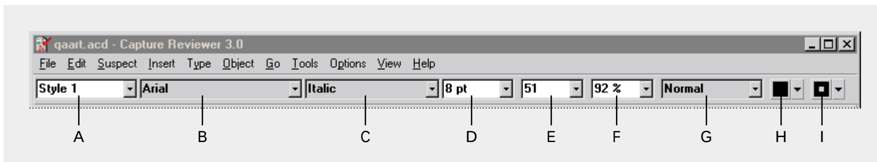
• To delete the character to the right of the insertion point, press Del.

To undo a text deletion:

Choose Edit > Undo.

Changing text formatting

You can change the font family, font face, font size, letterspacing, horizontal scale, position, and fill color of selected text. You can set these text properties separately or as a group. (You define the groups—*styles*—yourself.) In either case, you make the changes by choosing from menus on the format bar.



A. *Style* ***B.*** *Font family* ***C.*** *Font face* ***D.*** *Font size* ***E.*** *Letterspacing* ***F.*** *Horizontal scale* ***G.*** *Position* ***H.*** *Fill color* ***I.*** *Stroke color*

You can also copy text formatting from one character or group of characters to another.

To change character formats:

1 Use the type tool (T) to select the characters whose format you want to change.

2 Do one or more of the following:

• Choose a *font family* from the font family pop-up menu.

• Choose a *font face* from the font face pop-up menu. The choice of font faces depends on the font family chosen. For example, Helvetica, but not Times, has a Narrow font face. Similarly, the effect of changing font weight or choosing Italic depends on the font family. For example, choosing Italic with Helvetica changes the selected characters to Oblique.

• Choose a *point size* from the font size pop-up menu or enter the size in the size box. The maximum size allowed is 360 points.

• Choose a number from the *letterspace* pop-up menu. The numbers represent increases or decreases in the space between letters in thousandths of an em. A positive number increases the space. A negative number decreases the space.

• Choose a percentage from the *horizontal scale* pop-up menu. The percentages represent proportions between the height and width of the type relative to the original height and width. Unscaled characters have a value of 100%. Characters narrower than the original width have values below 100%.

• Choose a *position* on, above, or below the baseline (Normal, Superscript, Subscript) from the position pop-up menu. You can also choose Small Capital from the menu.

• Click the box to the left of the fill color pop-up menu to choose the *color* in it, or choose another color from the menu. Alternatively, choose Other from the menu, click a basic or custom color, and click OK. For information on custom colors, see “Creating custom colors” on page 116.

• Choose a *style* from the style pop-up menu. (Alternatively, choose Type > Styles, define a new style or modify an existing one, and choose that style. For information, see “Defining text styles” on page 103.)

To underline text:

1 Use the type tool (T) to select the characters you want to underline.

2 Choose Type > Underline.

To copy text formatting:

1 Select a sample of the text with the format you want to copy. The font, position, size, style, and color of the text—all its properties—will be copied.

2 Choose Type > Copy Format.

3 Select the text you want to format.

4 Choose Type > Paste Format.

Defining text styles

A text style is a set of text properties you define as a group—font family, font face, font size, letterspacing, and horizontal scale.

To define a new text style:

1 Select some text with the appearance you want for the style.

2 Click the style name in the format bar to select it.

3 Type a name for the new style over the selected name.

To edit a text style:

1 Choose Type > Styles.

2 Select the style to edit.

3 Make style changes as necessary.

4 Click OK.

To change the default text style:

1 Make sure that no text is selected. With no text selected, the current default style is the style displayed in the format bar.

2 Choose another style from the styles pop-up menu.

Using symbol characters as insertions or replacements

Symbol characters such as copyright (©) are often small and difficult to recognize. When Capture fails to recognize symbol characters, you can use one of two techniques to insert them into the converted ACD document. The easiest way to enter symbol characters is to select the symbol character you want to insert from a menu of symbol characters. However, you can also enter the three- or four-digit code for any special character, including symbol characters, in any font.

To use a symbol or special character as an insertion or replacement:

1 Use the type tool (T) to do one of the following:

- Position the insertion point.
- Select text to replace.

2 Do one of the following:

- Choose one of the symbol characters from the Insert menu.
- Choose Insert > Other Symbol, select the symbol character from the scrolling list, and click Insert.
- Hold down the Alt key, and enter the three- or four-digit code for the character.

3 If necessary, select the character, and change its format (font, size, or position, for example). Symbol characters are often best treated as superscripts.

You can use the Windows accessory Character Map to find the four-digit code for any special character in any font installed on your computer. See Windows Help for Character Map instructions.

Restoring images misrecognized as text

If a bitmap image is misrecognized as text, the text will probably be highlighted as a line or several lines of suspect words. In this case, you can't correct the error by editing the text. You have to convert the suspect text line back to the original bitmap image.

In other cases, an image may be misrecognized as a word but not considered suspect, perhaps because the word confidence level is too low to register on it. Or a suspect word needs to be displayed as a bitmap because it is in an unavailable font, but you have corrected it so that it can be searched for and need to restore its suspect status. In either case, you simply mark the word as suspect and then refuse to accept it.

To convert a text line to the original bitmap image:

1 Do one of the following:

- Use the selection tool ( ) to select the text line.
- Use the type tool (T) to select all the text in the line.

2 Choose Object > Change to Image.

Object > Change to Image also works with colored rectangles. See "Restoring and editing source images" on page 115.

To mark an unhighlighted word as a suspect:

1 Select the word.

2 Choose Suspect > Mark as Suspect.

Adding Reviewer suspects to dictionaries

The Capture Image step that produces the pages you work with in Reviewer uses at least one standard language dictionary for OCR purposes. It can also use user dictionaries for specialized word lists. Reviewer identifies a word as a Dictionary suspect only if it doesn't appear in any of these standard or user dictionaries.

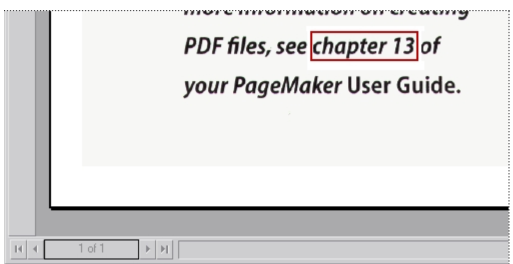
You can edit these user dictionaries while you are working in Reviewer. Alternatively, Reviewer provides a quick way of adding suspect words to user dictionaries. As well as adding to the dictionary for future use, the quick method has the immediate effect of accepting the suspect word and all other instances of the suspect in the document. For information, see "Working with user dictionaries" on page 122.

Note: *If the words you want to add are Dictionary suspects that will be useful only for the document you are processing, not for future documents, don't bother to add them to a user dictionary. You get the same effect of immediately accepting all instances of a Dictionary suspect in the document just by accepting one of them.*

About Reviewer's layout-recognition objects

The OCR step Capture Image divides each image it processes into rectangular objects of four types: text lines, graphic lines, graphic rectangles, and source-image rectangles. *Text lines* are regions that Capture Image recognizes as continuous lines of text. *Graphic lines and rectangles* are straight lines and solid-color rectangles that Capture Image recognizes and draws, preserving their original color. S*ource-image rectangles* are regions of the original bitmap displayed as is because Capture Image recognizes no part of them as text lines, graphic lines, or graphic rectangles.

In addition, Capture Image recognizes text that seems to refer to other locations as links, and superimposes rectangular link objects on this text. Link objects are normally invisible, but they can be displayed (as red rectangles) for editing and navigating.



When you use Reviewer to correct a page containing any of these types of objects, you have a chance to edit the objects. You can also create new text lines, graphic lines, graphic rectangles, and links, and insert new bitmap pictures, on the page.

The following sections describe some common recognition problems that you can solve by creating or editing these objects. The problems are grouped by object type, but note that they sometimes involve the interplay of different types, for example, the misrecognition of one type as another type.

About text lines

All of the word suspects you check and correct occur in text lines. With a simple layout and a good scan, the lines will be recognized correctly and you won't need to edit them.

In more difficult conditions, however, Capture Image may create text lines that span columns of text, combine two text lines with slightly different baselines, fail to recognize a region of text as text, and so on. To correct errors of this kind, you can create, delete, resize, split, merge, and change the order of text lines.

When you create new text lines, you can insert text in them. You can control the font, position, point size, style, and color of the text.

About links

Capture Image recognizes only linked text, not linked images. If you want linked images such as image maps, you need to create the links yourself.

Capture Image guesses the location of a link in text, inferring from clues that the text refers to another location. Obviously, its guesses are sometimes wrong. For example, it may fail to identify a link because it contains no visible page number or URL, or identify a number as a page number link when it is only a number. In cases like these, you can provide the missed link or delete the false link.

You might also use the Reviewer pass to update a URL link or redraw the link rectangle for easier hits.

About colored graphic lines and rectangles

With text containing rules or frames, you can often improve on the bitmap by using Reviewer's drawing tools to draw the line or rectangle yourself. You draw a frame by drawing a rectangle and specifying a fill color of "None."

When the rules or frames are in a second color, you may be able to save time and disk space by scanning in black and white and then drawing the colored objects yourself.

About source-image rectangles

When Capture Image identifies an image region correctly, the original image is displayed as is and no editing of it is necessary. When Capture Image incorrectly identifies an image region as text, however, you need to restore the original image. You do this by defining the region with a source-image rectangle.

You can also use a source-image rectangle to display the original image of a text region that you have no need to treat as text. Using it has the same effect as using Zone Tool before the Capture Image step to define the region as an image zone.

Working with Reviewer objects

You have a number of options for viewing objects, selecting and deselecting them, and deleting and undeleting them. You can align groups of objects and merge objects of the same type.

Choosing a viewing mode

In Reviewer's default viewing mode (Normal), you view both text and graphics, but you can also view text only, graphics only, or a PDF preview. You can edit only what appears on-screen.

Graphics Only is useful, for example, for quickly identifying text regions or extraneous marks incorrectly recognized as graphics. For workflows including a Convert Document to PDF step, PDF Preview shows what the page you are reviewing would look like, in its current state, as a PDF Formatted Text & Graphics document with suspects shown as bitmaps.

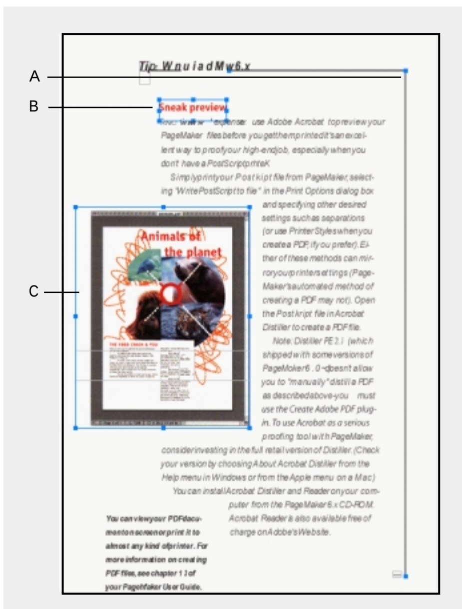
To choose a viewing mode:

Do one of the following:

- To display text only, choose View > Text Only.
- To display graphics only, choose View > Graphics Only.
- To display both text and graphics, choose View > Normal.
- To display a PDF preview, choose View > PDF Preview.

Selecting objects

You have to select an object to edit it. All selected objects have anchor points. Graphic rectangles have eight points. Text and graphic lines have two points, one at each end. Selected graphic rectangles have a blue border.



A. *Graphic line* ***B.*** *Text line* ***C.*** *Graphic rectangle*

To select a single object:

Select the selection tool (T), and do one of the following:

- To select a text line, click a character in the line. A selected text line is underlined and has *triangular* anchor points.
- To select a graphic line, click either end point of the line. Alternatively, select an end point by dragging a marquee (rectangle) to enclose it. A selected graphic line has *square* anchor points.
- To select a color-filled rectangle or source-image rectangle, click the image.

To select multiple objects:

Do one of the following:

• To select a cluster of objects, select the selection tool (T), and drag a marquee (box) around the objects.

• To select an object without deselecting other objects, hold down the Shift or Control key while selecting it with the selection tool.

• To select all objects, select the selection tool, and choose Select > Select All.

To deselect an object or objects:

Do one of the following:

• To deselect an object without deselecting or selecting other objects, select the selection tool (T), hold down the Control key, and click the object.

• To deselect all selected objects and select a previously unselected object, select the selection tool, and click the unselected object.

• To deselect all objects, select the selection tool, and either click an empty area in the page or choose Select > Deselect All.

Aligning objects

You can align objects with the top, bottom, left, or right edge of an object you want to align them with.

To align objects:

1 Select the objects you want to align.

2 Select the object you want to align them with.

3 Choose Object > Align.

4 Choose Left, Right, Top, or Bottom from the Align submenu.

Merging objects of the same type

You can merge any two or more objects of the same type.

With text line objects, merging is a way to re-create single text lines incorrectly recognized as two or more lines. In workflows with an Export to PDF or Export to HTML step, merging has no effect on the appearance of the PDF or HTML output, but it may make the results of selecting or copying text more coherent. In workflows with an Export to TXT step, merging can ensure the correct order of text in the TXT file.

To merge objects of the same type:

1 Select the objects.

2 Choose Object > Merge Objects.

Deleting objects

You can delete objects and undelete the most recent deletion.

To delete an object or objects:

1 Select the object or objects.

2 Do one of the following:

- Press Backspace.
- Press Delete.
- Choose Edit > Delete.

To undo the last deletion of an object or objects:

Choose Edit > Undo.

Creating and editing text lines

You can create, resize, merge, split, and reorder text lines. For information on viewing, selecting, and deleting text lines, see “Working with Reviewer objects” on page 106. For information on merging text lines, see “Merging objects of the same type” on page 108.

Creating a text line

If you want to enter text somewhere on the page where there is no text line, you need to create the line first.

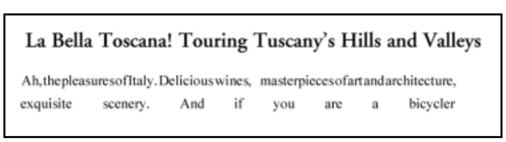
To create a text line:

1 Select the type tool (T).

2 Drag to draw out the baseline for the text line. When you finish dragging, the word *text* appears on the baseline in Helvetica. The text is the height provided for text by the default text style.

3 Double-click the word *text* to select it.

4 Type the new text. As you type, Reviewer keeps the line fully justified, so new words appear at the right side of the text line. (If you type too few characters, they are spaced too far apart. If you type too many, they are crowded.)



5 If necessary, adjust the length of the text line to the text. For information, see “Changing the size of a text line” on page 110.

6 If necessary, change the formatting of the text. For information, see “Changing text formatting” on page 102.

You can also enter text in a text line by pasting from the Clipboard, either from the current document in Reviewer or from a document in another application.

To copy and paste text:

1 Do one of the following:

- Copy text from another application.
- Using the Reviewer type tool (T), select the text, and choose Edit > Copy.

2 Using the Reviewer type tool, click in the area you would like to paste the text, and choose Edit > Paste.

Changing the size of a text line

When you change the length of a line, the text is justified for the new line length.

For information on creating text lines, see "Creating a text line" on page 109.

To change the length of a text line:

1 Select the line.

2 Position the cursor over one of the line's anchor points. The pointer changes into a directional arrow ( ).

3 Drag the anchor point left or right to change the length of the line.

Splitting text lines

You can split text lines that have been incorrectly recognized as a single line. In workflows with an Export to PDF or Export to HTML step, splitting has no effect on the appearance of the PDF or HTML output, but it may improve the results of selecting or copying text by eliminating breaks in its logical order.

Splitting text lines creates more possibilities for reordering the lines. In workflows with an Export to TXT step, where text in the output file is in the order of the text lines, this means more control over the appearance of the output.

To split text lines:

1 Select the type tool (T).

2 Click the point in the text line where you would like to split the line.

3 Choose Type > Split Text Line.

Reordering text lines

The text lines recognized in the Capture Image step are given a numerical order that can sometimes be wrong, for example, on a multicolumn page with several sections, each designed to be read in left-right order. You can correct the order in Reviewer.

In workflows with an Export to PDF or Export to HTML step, reordering text lines has no effect on the appearance of the PDF or HTML output, but it may improve the results of selecting or copying text by restoring the logical order of the text. In workflows with an Export to TXT step, text in the output file is in the order of the text lines. Whatever the export type, correctly ordering text lines makes documents more legible to visually impaired readers who are hearing rather than seeing them.

To reorder text lines within a page:

1 Select the text order tool ( ). Numbered squares appear in each text line to show the order of text lines on the page.

2 Do one of the following:

• If the line that should be first is followed by a line that shouldn't follow it, click the line that should be first.

• If the line that should be first is followed by one or more lines that should follow it in order, drag through this group of lines.

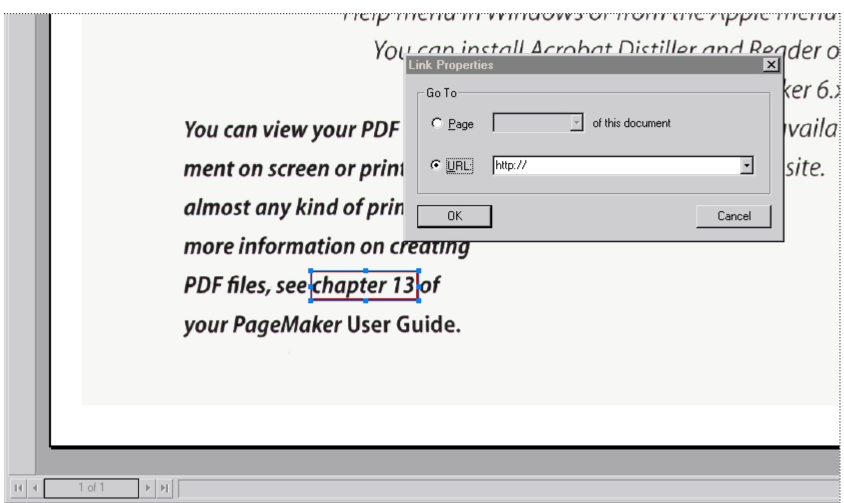
3 Continue clicking individual lines or dragging through well-ordered groups of lines to establish the correct order.

4 Deselect the text order tool.

Creating and editing links

Links are created during the Capture Image step, and you can create, delete, and edit them yourself with Reviewer. Links can be URLs or references to pages. The Capture Image step recognizes references to both simple page labels such as *35* and *viii* and complex ones such as *C-37*.

Links are normally invisible. You display them (as red rectangles) to edit and navigate them.



You select, draw, and redraw link rectangles as you do graphic and source-image rectangles.

To display existing links:

Select the link tool ( ).

To create a new link:

1 Use the link tool ( ) to draw a rectangle defining the area of the link.

2 Specify the link's destination.

To specify or change a link's destination:

1 Do one of the following:

• Double-click the link.

• Select the link, and choose Edit > Properties.

2 Do one of the following:

• Select Page, and select a page label from the pop-up menu.

• Select URL, and enter a URL string. The pop-up menu provides protocols for the string, for example, http://.

3 Click OK.

To jump to the destination of a link:

1 Remember where the link is. It won't be displayed during this procedure.

2 Select the hand tool ( ).

3 Move the hand over the link. When it is over the link, it changes to a pointing finger ( ).

4 Click.

Creating and editing graphic objects

You can add colored lines and rectangles to a page, insert pictures in it, and layer these graphic objects as you wish.

For information on viewing, selecting, and deleting graphic objects, see "Working with Reviewer objects" on page 106.

Adding colored lines to a page

When you add a line, you can control its position, direction, length, color, width, and cap style (round or square).

If the line overlaps text, the text appears on top. If it overlaps an image or rectangle, the line can appear either on top or beneath. For information, see "Placing objects on top of other objects" on page 115.

To create a line:

1 Select the line tool ( ), and position the cursor over a nontext part of the page.

2 Drag to draw a line.

To change the line cap style to round or square:

1 Select the line.

2 Do one of the following:

- Choose Object > Line Cap Style > Round.
- Choose Object > Line Cap Style > Square.

To change the direction or length of a line:

1 Select the line.

2 Position the cursor over one of the line’s anchor points. The pointer changes into a directional arrow ( ).

3 Drag the anchor point in any direction to change the direction or length of the line.

To change the color of one or more lines:

1 Select the line or lines.

2 Do one of the following:

- To change to the color displayed in the box to the left of the stroke color pop-up menu, click the box.
- To change to another color, choose it from the stroke color pop-up menu.

Alternatively, choose Other from the menu, click a basic or custom color, and click OK. For information on custom colors, see “Creating custom colors” on page 116.

To change the weight of one or more lines:

1 Select the line or lines.

2 Choose Object > Stroke Weight.

3 Select a weight from the submenu.

Note: *If you change the stroke weight without selecting any lines, the change will apply to future lines you create. That is, you will be changing the default stroke weight.*

Adding colored rectangles to a page

When you add a rectangle, you can control its position, proportions, size, and color.

If the rectangle overlaps text, the text appears on top. If it overlaps another rectangle, a graphic line, or a source image, it can appear either on top or beneath. For information, see “Placing objects on top of other objects” on page 115.

To create a rectangle:

1 Select the rectangle tool ( ), and position the cursor over a nontext part of the page.

2 Drag to draw the outline of the rectangle.

To change the size of a rectangle:

1 Select the rectangle.

2 Position the pointer over an anchor point. The pointer changes into a directional arrow ( ).

3 Drag the anchor point in any direction to resize the rectangle.

To move a rectangle:

1 Select the rectangle.

2 Position the cursor over the rectangle. The pointer changes into a four-directional arrow ( ).

3 Drag the rectangle to its new location on the page.

To change the weight of a rectangle's border:

1 Select the rectangle.

2 Choose Object > Stroke Weight.

3 Select a weight from the submenu.

Note: *If you change the stroke weight without selecting any rectangles, the change will apply to the borders of future rectangles you create. That is, you will be changing the default stroke weight.*

To change the color of a rectangle's border:

1 Select the rectangle or rectangles.

2 Do one of the following:

• To change to the color displayed in the box to the left of the stroke color pop-up menu, click the box.

• To change to another color, choose it from the stroke color pop-up menu. One choice is None—no border.

Alternatively, choose Other from the menu, click a basic or custom color, and click OK. For information on custom colors, see "Creating custom colors" on page 116.

To change the interior (fill) color of a rectangle:

1 Select the rectangle or rectangles.

2 Do one of the following:

• To change to the color displayed in the box to the left of the fill color pop-up menu, click the box.

• To change to another color, choose it from the fill color pop-up menu. One choice is None—no fill.

Alternatively, choose Other from the menu, click a basic or custom color, and click OK. For information on custom colors, see "Creating custom colors" on page 116.

You can create a frame for text by choosing a stroke (border) color and None for the fill.

Inserting pictures

You can insert TIF, PDF Image, BMP, and PCX files as pictures. When you insert a picture, it becomes an object on the page that you can work with in the same way you would work with a source image.

To insert a picture:

1 Choose Insert > Picture.

2 Select the image file, and click OK. The picture is positioned in the upper left corner.

3 If the picture will replace a source image, delete the bitmap.

4 Move the picture in the correct location on the page.

Restoring and editing source images

You can restore the source image of any region on a page. The region can be one you have already described by drawing a colored rectangle, or one you draw now for the purpose of displaying its source image.

You can also provide any source image with a searchable keyword.

To change a rectangle to its source image:

1 Select the rectangle.

2 Choose Object > Change to Image.

Object > Change to Image also works with text lines. See “Restoring images misrecognized as text” on page 104.

To draw a region’s source image:

1 Select the source image tool ().

2 Draw a rectangle over the region.

To provide a hidden keyword for a source-image rectangle:

1 Do one of the following:

- Double-click the rectangle.
- Select the rectangle, and choose Edit > Properties.

2 Enter the keyword.

3 Click OK.

Placing objects on top of other objects

When a source-image rectangle and a graphic object, or two graphic objects, overlap, you can position either object on top of the other. (Text objects always appear on top of source-image and graphic objects.)

To move one object on top of another:

1 Select one of the objects you want to move.

2 Choose Object > Push to Bottom or Object > Bring to Top as appropriate.

Creating custom colors

You choose colors for text and graphic objects from the stroke color and fill color menus in the format bar. These menus list both basic colors and any custom colors you create yourself.

To create a custom color:

1 Do one of the following:

- Choose Other from the stroke color pop-up menu.
- Choose Other from the fill color pop-up menu.

2 Click Define Custom Colors.

3 Define the color. For instructions, see What's This? help for the dialog box.

4 Click Add to Custom Colors.

5 Click OK.

Reviewer shortcuts

The shortcuts listed here have no equivalent commands. For the shortcut to a command on a menu, see the menu.

Moving around	
Result	**Action**
Move left one character	Left Arrow
Move right one character	Right Arrow
Move up one line	Up Arrow
Move down one line	Down Arrow
Move to previous word	Ctrl+Left Arrow
Move to next word	Ctrl+Right Arrow
Scroll view up one line	Ctrl+Up Arrow
Scroll view down one line	Ctrl+Down Arrow
Move to beginning of text line	Home
Move to end of text line	End
Move to beginning of document	Ctrl+Home
Move to end of document	Ctrl+End
Move up one screen-full	Page Up
Move down one screen-full	Page Down
Delete to left one character	Backspace

Moving around	
Result	**Action**
Delete to right one character	Del
Move and select text	Shift+<navigation keys>

Nudging objects	
Result	**Action**
Move selected objects by 1 pt	Arrow keys
Move selected objects by 10 pts	Shift+arrow keys

Modifying the selection tool	
Result	**Action**
Select this object only	Click
Toggle selection of object	Ctrl+click
Add to selected objects	Shift+click
Select cluster of objects	Alt+click
Move or size selected objects	Drag
Perform constrained move or size	Shift+drag

Chapter 11

Using metadata, suspects, and dictionaries

You can use the properties dialog boxes for several steps to provide document information (metadata), suspect settings, and dictionary entries. Also, you can provide metadata in Zone Tool and dictionary entries in QuickFix and Reviewer.

Providing document information (metadata)

You can provide metadata for documents processed in a workflow by configuring any one of several steps in the workflow: Archive File, Export to HTML, Export to PDF, or Zone Image. With Zone Image, you can also provide metadata for a document or image from within the associated Zone Tool application, while you are working on the document or image. The same holds true for the Reviewer application associated with the Review Document step.

Whether you are configuring a step or using an application, you use the same Edit Document Info dialog box. The default setting of Edit Document Info is “no documentation information.”

About document information options

You can provide standard document information values and ODMA-specific values for metadata keys. The standard values are Title, Subject, Author, and Keyword. The ODMA-specific attributes are Location and Type.

In addition, you can provide new keys for attributes used by an ODMA-compliant DMS and new general-use keys.

ODMA-specific and DMS-specific attributes apply only to workflows that include a destination step adapted to them. (Of the steps delivered with Acrobat Capture, only Archive File is so adapted. For information, see “Archive File” on page 8.) However, which step you configure in order to provide metadata values of these two types, or any types, is usually simply a matter of convenience. For example, you could provide ODMA-specific values while configuring an Export to PDF step in a workflow that contained an Archive File step.

A general-use key used in a step needs to use key names recognized by the software that performs the step. For a complete list of the keys recognized by the step templates delivered with Acrobat Capture, see “Acrobat Capture SDK Overview” at http://partners.adobe.com/asn/developer/technotes.html.

Whatever step you configure, the values you provide in it will be overridden by values provided in a JOB file or during a Zone Tool or Reviewer session. (Zone Tool and Reviewer values override JOB file values.) JOB files are used in submissions to watched folders. For more information on JOB files, see “Providing metadata with a JOB file” on page 42. For information on Zone Tool and Reviewer, see “Zoning Images for Capture” on page 65 and “Reviewing Captured Text and Graphics” on page 93.

Some uses of metadata specific to other steps are as follows.

Mail File If you specify a MailTo key in the Document Info dialog box or in a JOB file, this key's value will override the Recipients field in the Mail File Properties dialog box. For information, see “Configuring Mail File” on page 24 and “Providing metadata with a JOB file” on page 42.

Review Document The metadata values you provide will typically differ from document to document. One way to allow for this is to submit a single document at a time and customize processing for each submission. Another way is to include a Review Document step in the workflow and provide the values separately for each document during that step. In this case, you would use the Reviewer application to provide the values rather than configuring another step with them.

Zone Image If you are processing a set of pages that have keyword values in a fixed location, you can use a default zone file to extract the values automatically. For information, see “Using template and default zone files in workflows” on page 75. If you also provide a keyword value in the default zone file by configuring the Zone Image step, any extracted value is appended to the value you provide. For example, if you were processing numbered claim forms whose numbers were being extracted automatically and if you provided the keyword “claim,” form 0173 could be searched by “claim AND 0173.”

Also, you can use the Zone Tool application to provide separate document information values for each document submitted to a workflow. (In this case, you use the Zone Tool application to provide the values rather than configuring the Zone Image step.)

Editing document information

For editing document information, you use the same methods whether you are configuring a step or working in Zone Tool or Reviewer.

To edit metadata keys:

1 Do one of the following:

- Click Document Info in the properties dialog for the step you are configuring.
- In Reviewer or Zone Tool, choose File > Document Info.

2 Do one of the following:

- Provide values for keys.
- Create new keys.
- Delete or rename keys you have created.

3 Click OK.

To provide values for keys:

1 Select a key that requires a new value.

2 Type a new value for the key.

3 Click Set.

4 Repeat the process for the next key that requires a new value.

To create new keys:

1 Click New Key.

2 Provide a key name. The name can contain any characters except = (equal sign).

3 Specify a key type. The key can be for use by an ODMA-compliant DMS or for general use.

4 Repeat the process for the next key you want to create.

To delete a key you have created:

1 Select the key.

2 Click Delete Key.

3 Repeat the process for the next key you want to delete.

To rename a key you have created:

1 Select the key.

2 Click Rename Key.

3 Provide a new key name. The name can contain any characters except = (equal sign).

4 Repeat the process for the next key you want to rename.

Working with suspects

Suspects are words that may have been incorrectly recognized during OCR.

About suspects

In workflows, the data required to define suspects is created during the Capture Image OCR step. During the step, each word captured is given a percentage score measuring confidence that the characters in it have been recognized correctly. Low confidence may reflect broken type, an unusual font, or a poor scan.

Also during the step, a word is flagged if its font is uncertain, if it isn't found in a dictionary used with the step, or if it contains mixed alphabetical and numeric characters or numeric characters only. (Alphanumeric and numeric "words" are flagged because they can't be checked against a dictionary.)

Potential and actual suspects Low-confidence words and flagged words are potential suspects. You define the actual suspects by specifying a confidence threshold and deciding what flags to recognize.

Uses of suspects Suspects get special treatment in two conversion steps and two OCR correction steps:

• In Export to HTML steps and in Export to PDF steps with page content set to Formatted Text & Graphics, suspect words are replaced in the HTML or PDF document by the original bitmap they were captured from.

• In QuickFix Page and Review Document steps, suspect words are called out for checking in the QuickFix or Reviewer application. Words accepted as is, or corrected and then accepted, lose their suspect status.

At any of these points, you can establish a confidence threshold below which you consider a word suspect, and you can decide which other flags (for example, uncertain font) also make a word suspect. By default, the threshold is 95% for all steps. The other flag defaults are as follows: none for Export to HTML and Export to PDF steps, all for QuickFix Page steps, and all but Numeric for Review Document steps.

Defining suspects

The default suspect criteria and confidence percentage are the same for Export to HTML, Export to PDF, QuickFix Page, and Review Document. If you use a QuickFix Page or Review Document step in a workflow with an Export step and change the defaults for the Export step, you should normally make corresponding changes in the other step. In particular, the QuickFix Page or Review Document step should have settings at least as conservative—that is, the step should define at least as many suspects—as the Export step.

Similarly, if a workflow has multiple export steps with different suspect settings, QuickFix Page or Review Document should be at least as conservative as the most conservative export step.

To define suspects:

1 Do one of the following:

- (Export to HTML or Export to PDF) Open the properties dialog box, and bring the Suspects panel to the front.
- (QuickFix Page or Review Document) Open the properties dialog box.

2 Select or deselect any of the suspect criteria, as necessary. (The effect of deselecting all criteria, including confidence, would be to prevent any checking in QuickFix and any bitmap substitution to HTML or PDF.)

3 Change the confidence percentage, as necessary. (The effect of decreasing the percentage would be to decrease the number of suspects. If you want to use the confidence criterion, be sure it is selected (checked) as well.)

4 Click OK.

Working with user dictionaries

You can create your own dictionaries and add words to them as you are configuring the Capture Page step, or while you are working in the QuickFix or Reviewer application. You can add words one at a time or by importing word lists. You use the same Edit Dictionaries dialog box in all cases.

There are shortcuts for adding Dictionary suspects to dictionaries from QuickFix or Reviewer.

About user dictionaries

You create your own dictionaries (user dictionaries) to supplement Acrobat Capture's standard dictionaries. During a Capture Image step, Acrobat Capture uses one or more standard dictionaries and any user dictionaries you select for word recognition. If a recognized word does not appear in any of the dictionaries used with the workflow containing the step, the word is tagged as a dictionary suspect. Dictionary suspects can be replaced by bitmaps of the original word in Export to PDF and Export to HTML steps. If the workflow has a QuickFix Page or Review Document step before the export step, dictionary suspects are displayed so that you can check and correct them.

You can select any user dictionary, or any set of user dictionaries, for a particular workflow. Because you can combine dictionaries freely in this way, it may make sense to create a number of special-purpose user dictionaries rather than a single comprehensive one. For example, you could create one dictionary for acronyms, one for personal names, or one for a particular client or job category such as part numbers.

Creating and editing user dictionaries

For creating and editing user dictionaries, you use the same methods whether you are configuring the Capture Page step or working in QuickFix or Reviewer.

To open the Edit Dictionaries dialog box:

Do one of the following:

- In QuickFix, choose Options > Edit Dictionaries.
- In Reviewer, choose Options > Edit Dictionaries.
- In the Capture Image Properties dialog box, click New opposite the Dictionaries list. This also creates a new user dictionary.
- In the Capture Image Properties dialog box, select a user dictionary, and click Edit.

To create a new user dictionary:

1 Do one of the following:

- Click New in the Edit Dictionaries dialog box.
- Click New in the Capture Image Properties dialog box. This also opens the Edit Dictionaries dialog box.

2 Give the new dictionary a name. (Change its name from “Untitled.”)

3 Edit the dictionary.

To edit a user dictionary:

1 Do one of the following:

- If the Edit Dictionaries dialog box is open, select the dictionary. (Take this option if you are in the process of adding a new dictionary.)
- If the Edit Dictionaries dialog box is not open, select the dictionary, and click Edit.

2 Rename or delete the dictionary, as necessary.

3 Add, remove, modify, or import words, as necessary.

4 Click Close.

To rename or delete a user dictionary:

1 Make sure that no workflows with steps that use user dictionaries are running. (These steps include Capture Image, QuickFix Page, and Review Document.)

2 Select the dictionary.

3 Do one of the following:

- Click Delete.
- Click Rename, and give the dictionary a new name.

To add a word to a user dictionary:

1 Type the word in the Word box.

2 Click Add.

Adding a word from QuickFix or Reviewer has the further effect of accepting all instances of the word that are Dictionary suspects on the current page and all further pages of the document.

To remove or modify a word in a user dictionary:

1 Select the word.

2 Do one of the following:

- Click Remove.
- Modify the word in the word box, and click Modify.

You can import words from any text file (.txt or .dic) that has one word per line. Importing merges the contents of the imported file with the current contents of the user dictionary.

To import an existing dictionary to a user dictionary:

1 Click Import.

2 Navigate to the file, and click Open.

You can import words from any text file (.txt or .dic) that has one word per line. Importing merges the contents of the imported file with the current contents of the user dictionary.

To export a user dictionary:

1 Click Export.

2 Change the file type (.dic by default), if necessary.

3 Provide a filename.

4 Click Save.

Adding suspect words to user dictionaries from QuickFix and Reviewer

QuickFix and Reviewer provide a quick way of adding suspect words to user dictionaries. As well as adding to the dictionary for future use, the quick method has the further effect, with Dictionary suspects, of accepting the suspect word and all other instances of the suspect in the document.

Note: *If the words you want to add are Dictionary suspects that will be useful only for the document you are processing, not for future documents, don't bother to add them to a user dictionary. Just accepting a Dictionary suspect will accept all other instances of that suspect in the document.*

To add a suspect word to a user dictionary:

1 Do one of the following:

- In QuickFix, select the row containing the Dictionary suspect.
- In Reviewer, select the Dictionary suspect.

2 Choose Add to Dictionary from the QuickFix Edit menu or the Reviewer Suspect menu.

3 Select a dictionary from the list. (The list is the same as the Dictionaries list in the Capture Image Properties dialog box.)

4 Click OK. This accepts the word if it is a Dictionary suspect, and accepts all instances of the same word that are Dictionary suspects. (It accepts them both on this page and on all further pages of the document—on all pages not yet processed in QuickFix or Reviewer.) If you correct the suspect before accepting it, all instances of the same incorrect word are *not* corrected and accepted as corrected.

To add a suspect word to the user dictionary last used at your station:

1 Do one of the following:

- In QuickFix, select the row containing the Dictionary suspect.
- In Reviewer, select the Dictionary suspect.

2 Choose Add to Last from the QuickFix Edit menu or the Reviewer Suspect menu. (When you open the menu to choose the command, “Last” is replaced by the name of the last used user dictionary.)

Once you have added a suspect word to a user dictionary in QuickFix or Reviewer, you can use a simpler method for adding other words to that dictionary. The method works only if you haven’t used another user dictionary in the meantime.

Index

T

U

V

W